

(19)日本国特許庁 (JP)

(12)公開特許公報 (A)

(11)特許出願公開番号

特開平11-134791

(43)公開日 平成11年(1999)5月21日

(51)Int.Cl. <sup>9</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G11B 20/10			G11B 20/10	B
	321			321 Z
20/12	102		20/12	102
// G09B 5/04			G09B 5/04	

審査請求 有 請求項の数6 O L (全20頁)

(21)出願番号 特願平9-294384

(22)出願日 平成9年(1997)10月27日

(71)出願人 000104179

カナース・データー株式会社

東京都千代田区東神田1丁目10番7号

(72)発明者 関口 博司

東京都千代田区東神田1-10-7 カナース・データー株式会社内

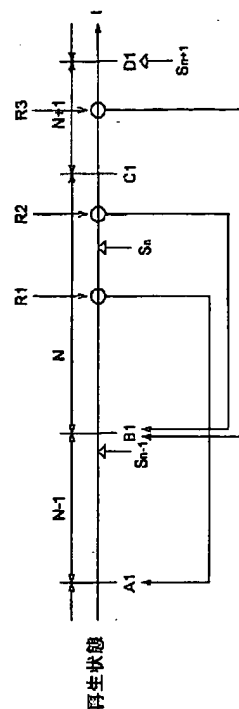
(74)代理人 弁理士 長谷川 芳樹 (外3名)

(54)【発明の名称】 音声記録媒体、その再生方法、及び音声再生中の割込み処理方法

(57)【要約】

【課題】 長さの異なる複数のセンテンスから構成された一連の音声情報を再生する場合であっても、操作者の指示遅れも考慮されたフレキシブルな音声再生を可能にする構造を備えた音声記録媒体、その再生方法、及び割込み処理方法を提供する。

【解決手段】 音声情報とともに、該音声情報列を構成する可変長区画に関する情報として、記録媒体における該各区画( $\dots, N-1, N, N+1, \dots$ )を発音のまとまりと判断された前半部分と後半部分とに区切る境界位置( $S_{i-1}, S_i, S_{i+1}$ )を示す境界位置識別情報が少なくとも記録された音声記録媒体を用意し、所定の割込み要求のタイミング( $R_1, R_2, R_3$ )が再生中区画の境界位置( $S_{i-1}, S_i, S_{i+1}$ )の前か後かを判断し、該判断結果に基づいて割込み要求の対象区画を特定する。



## 【特許請求の範囲】

【請求項 1】 所定の音声再生手段で再生出力されるべき複数の単語列から構成された 1 又は 2 以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された音声情報ごとに可変長の区画に区分された音声情報列が記録された第 1 領域と、該音声情報列における各区画ごとに、該各区画の前半部分と後半部分との境界を示す境界位置識別情報が記録された第 2 領域とを、少なくとも備えた音声記録媒体の音声情報再生中に発生した、種々の再生方法を指示する割込み要求の処理を実行するための音声再生中の割込み処理方法において、前記区画のうち所定区画の再生中に発生した割込み要求に対し、該割込み要求が、該再生中であった区画の先頭位置から境界位置までの前半部分で発生したか、あるいは該再生中であった区画の境界位置から次区画の先頭位置までの後半部分で発生したかを判断し、前記割込み要求の発生タイミングに関する判断結果に基づいて、前記再生中であった区画の前半部分で発生したと判断された前記割込み要求が対象とする区画として、該再生中であった区画の前区画を特定するか、あるいは前記再生中であった区画の後半部分で発生したと判断された前記割込み要求が対象とする区画として、前記再生中であった区画自身を特定することを特徴とする音声再生中の割込み処理方法。

【請求項 2】 所定の記録媒体に記録された音声情報を再生する、音声記録媒体の再生方法において、所定の音声再生手段で再生出力されるべき複数の単語列から構成された 1 又は 2 以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された音声情報ごとに可変長の区画に区分された第 1 音声情報列が記録された第 1 領域と、該第 1 音声情報列と等価でかつ遅延した発音で構成された音声情報列であって、該第 1 音声情報列の各区画の音声情報と等価な音声情報ごとに可変長の区画に分割された第 2 音声情報列が記録された第 2 領域と、該第 1 及び第 2 音声情報列を該音声再生手段で切替えて再生すべく、切替可能な各区画を当該記録媒体における該各区画の記録位置で示す記録位置識別情報が記録された第 3 領域と、そして、該第 1 及び第 2 音声情報列における各区画ごとに、該各区画の前半部分と後半部分との境界を示す境界位置識別情報が記録された第 4 領域とを、少なくとも備えた音声記録媒体を用意し、前記区画のうち所定区画の再生中に発生した前記第 1 及び第 2 音声情報列間での切替え再生に関する第 1 の割込み要求に対し、該第 1 の割込み要求が、該再生中であった区画の先頭位置から境界位置までの前半部分で発生したか、あるいは該再生中であった区画の境界位置から次区画の先頭位置までの後半部分で発生したかを判断し、前記第 1 の割込み要求の発生タイミングに関する判断結果に基づいて、前記再生中であった区画の前半部分で発生したと判断された該第 1 の割込み要求が対象とする区

画として、該再生中であった区画の前区画を特定するか、あるいは前記再生中であった区画の後半部分で発生したと判断された前記第 1 の割込み要求が対象とする区画として、前記再生中であった区画自身を特定し、前記第 1 の割込み要求が対象とする区画として特定された区画に対応する区画であって、前記第 1 及び第 2 音声情報列のうち、該第 1 の割込み要求によって指示された音声情報列の区画の先頭から、音声再生を再開することを特徴とする、音声記録媒体の再生方法。

10 【請求項 3】 前記区画のうち所定区画の再生中に発生したリピート再生に関する第 2 の割込み要求に対し、該第 2 の割込み要求が、該再生中であった区画の先頭位置から境界位置までの前半部分で発生したか、あるいは該再生中であった区画の境界位置から次区画の先頭位置までの後半部分で発生したかを判断し、前記第 2 の割込み要求の発生タイミングに関する判断結果に基づいて、前記再生中であった区画の前半部分で発生したと判断された該第 2 割込み要求が対象とする区画として、前記再生中であった区画の前区画を特定するか、あるいは前記再生中であった区画の後半部分で発生したと判断された前記第 2 の割込み要求が対象とする区画として、前記再生中であった区画自身を特定し、前記第 1 及び第 2 音声情報列のうち、再生中であった区画を含む音声情報列であって、前記第 2 の割込み要求が対象とする区画として特定された区画の先頭から、音声再生を再開することを特徴とする、請求項 2 記載の音声記録媒体の再生方法。

【請求項 4】 所定の記録媒体に記録された音声情報を再生する、音声記録媒体の再生方法において、所定の音声再生手段で再生出力されるべき複数の単語列から構成された 1 又は 2 以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された音声情報ごとに可変長の区画に区分された音声情報列が記録された第 1 領域と、該音声情報列における各区画ごとに、該各区画の前半部分と後半部分との境界を示す境界位置識別情報が記録された第 2 領域とを、少なくとも備えた音声記録媒体を用意し、前記区画のうち所定区画の再生中に発生したリピート再生に関する割込み要求に対し、該割込み要求が、該再生中であった区画の先頭位置から境界位置までの前半部分で発生したか、あるいは該再生中であった区画の境界位置から次区画の先頭位置までの後半部分で発生したかを判断し、前記割込み要求の発生タイミングに関する判断結果に基づいて、前記再生中であった区画の前半部分で発生したと判断された前記割込み要求が対象とする区画として、該再生中であった区画の前区画を特定するか、あるいは前記再生中であった区画の後半部分で発生したと判断された前記割込み要求が対象とする区画として、前記再生中であった区画自身を特定し、

前記割込み要求が対象とする区画として特定された区画の先頭から、音声再生を再開することを特徴とする、音声記録媒体の再生方法。

【請求項 5】 所定の音声再生手段で再生出力されるべき複数の単語列から構成された 1 又は 2 以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された音声情報ごとに可変長の区画に区分された音声情報列が記録された第 1 領域と、

前記音声情報列における各区画ごとに、該各区画の前半部分と後半部分との境界を示す境界位置識別情報が記録された第 2 領域とを、少なくとも記録された音声記録媒体。

【請求項 6】 所定の音声再生手段で再生出力されるべき複数の単語列から構成された 1 又は 2 以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された音声情報ごとに可変長の区画に区分された第 1 音声情報列が記録された第 1 領域と、

前記第 1 音声情報列と等価でかつ遅延した発音で構成された音声情報列であって、該第 1 音声情報列の各区画の音声情報と等価な音声情報ごとに可変長の区画に分割された第 2 音声情報列が記録された第 2 領域と、

前記第 1 及び第 2 音声情報列を該音声再生手段で切換えて再生すべく、切換え可能な各区画を当該記録媒体における該各区画の記録位置で示す記録位置識別情報が記録された第 3 領域と、そして、

前記第 1 及び第 2 音声情報列における各区画ごとに、該各区画の前半部分と後半部分との境界を示す境界位置識別情報が記録された第 4 領域とを、少なくとも備えた音声記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、CD-ROM、DAT等の媒体に少なくとも音声情報を含む種々の情報が記録された音声記録媒体と、このような音声記録媒体にあらかじめ記録された音声情報を再生する再生方法と、音声再生中に発生する、種々の再生方法を指示する割込み要求を処理するための割込み処理方法に関するものである。

【0002】

【従来の技術】従来から、英会話等の語学の独習用、詩吟の練習用、法律の独習用、その他の目的のために、カセットテープ等の記録媒体に音声情報が記録された教材が種々提供されている。ここで、英会話の独習用の教材を例に説明すると、従来の主な記録媒体は、例えば一連の英語の発声（音声情報）が記録されたカセットテープ（又はレコード）であり、学習者はこのテープ教材とテキストとを組み合わせ使用していた。なお、このような教材には、初級用から上級用まで種々のレベルが用意されている。

【0003】また、日本国特許第 2 5 8 1 7 0 0 号に

は、複数の区画に区分された上級者学習用に適した音声情報列が記録された第 1 領域と、これら各区画に対応した等価な区画からなる初級者学習用に適した音声情報列が記録された第 2 領域と、該上級者学習用及び初級者学習用の各音声情報列の対応する各区画の関係を、これら音声情報列の各区画の記録媒体における記録位置で示す情報が記録された第 3 領域とを、少なくとも備えた CD-ROM等の情報記録媒体、及びこのような構造を備えた情報記録媒体の切替え再生等を含む再生方法が提案されている。

【0004】一方、特開平 5 - 2 2 4 5 8 1 号公報には、語学学習用 CD を利用した音声再生中のリピート機能を実現するための技術が提案されている。すなわち、この技術は通常のヒヤリング学習中における学習者からの指示遅れを感知する遅延感知機能に関するものであり、具体的には、あらかじめインデックスナンバーが付された 1 センテンス（文）ごとの音声情報列が順次記録された媒体を所定の再生装置を用いて再生している。例えば、このような再生装置にリピート再生ができる機能が設けられていた場合であっても、難解なセンテンスや、聞き取りにくい発音を含んだセンテンスを聞いた直後等では、リピート指示（学習者による機器操作）が遅れて次のセンテンスの再生が始まってしまい、学習者は既に聞き終わったセンテンスを聞くことができなくなるという不具合があった。そこで、この従来技術では、図 1 9 に示されたように、順次 A、B、C、D の各時点で再生対象となるセグメントのインデックスナンバーが順次切替えられることにより、再生対象となるセグメントの音声情報が再生されるが、同時に次の音声情報の再生開始から時間  $\Delta$  経過まで前のセグメントを示すインデックスナンバー（リピート再生する音声情報を示すインデックスナンバー）も所定の保持手段により保持される。換言すれば、この従来技術は再生対象であるセグメントのインデックスナンバーとリピート再生の対象であるべきセグメントのインデックスナンバーとを独立に保持し、これら各インデックスナンバーの切替えタイミングを所定時間  $\Delta$  だけずらすよう時間管理している。このように構成することにより、学習者からのリピート要求が発行された時点での保持されていたインデックスナンバーに対応するジャンプ信号（再生開始位置を指示する情報）を発生させ、学習者からのリピート要求の遅れに起因した問題（指示遅れ）を回避している。

【0005】

【発明が解決しようとする課題】以上のように、従来の音声再生方法では、リピート機能のような操作者（例えば、語学学習者）からの指示遅れをあらかじめ考慮し、操作者が意図したセンテンスからリピート再生が行えるよう、該操作者からの指示の遅延感知が行われている。

【0006】しかしながら、このような遅延感知機能は所定のセンテンスが順次規則的に記録された場合を想定

した技術である。したがって、例えば映画における出演者の会話、日常生活環境における会話等のように、長さの異なる複数のセンテンスから構成され、また、各センテンス（各会話者の音声情報）の間に、音声再生されていない状況、雑音のみが再生されている状況、音楽（BGM）のみが再生されている状況等のランダムに発生する無音声期間が長時間（例えば1秒以上）存在し得る一連の音声情報の再生では、時間管理による遅延感知機能では操作者である学習者が希望する動作を的確に実現するのは困難である。すなわち、このような状況下では操作者から意図したリピート対象センテンスと、実際にリピート再生されるセンテンスとが異なる場合が起こり、結果的に語学学習等では機器操作に学習者の注意が奪われ十分な学習効果が低減してしまう（集中力が散漫になる）という課題があった。

【0007】そこで、この発明は上述のような課題を解決するためになされたものであり、長さの異なる複数のセンテンスから構成された一連の音声情報を再生する場合であっても、的確に操作者が意図する音声情報の再生を可能にする構造を備えた音声記録媒体、その再生方法、及び音声再生中の割込み処理方法を提供することを目的としている。

【0008】

【課題を解決するための手段】この発明は、映画における出演者の会話、日常生活環境における会話等のように、長さの異なる複数のセンテンスから構成され、また、各センテンス（各会話者の音声情報）の間に、音声再生されていない状況、雑音のみが再生されている状況、音楽（BGM）のみが再生されている状況等のランダムに発生する無音声期間が存在し得る一連の音声情報の再生においても、操作者が意図する音声情報の再生を、指示遅れも考慮して的確に実現する技術に関するものである。

【0009】特に、再生対象である音声記録媒体には、所定の音声再生手段で再生出力されるべき複数の単語列から構成された1又は2以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された音声情報ごとに可変長の区画に区分された1種類又は2種類以上の音声情報列が少なくとも記録されている。なお、この明細書において、音声情報とは、実際に聞き取れる音（音声）を含む概念及び情報であり、音声情報列とは、実際にCD-ROM等の記録媒体に記録されたデータ自体を意味する。

【0010】具体的に、この発明に係る音声記録媒体（音声記録媒体に係る第1実施態様）は、所定の音声再生手段で再生出力されるべき複数の単語列から構成された1又は2以上の文に対応する音声情報列であって、発音の節目（センテンスの終りやセンテンス中の一息つける、発生上あるいは文法上の区切り）でそれぞれ分割された音声情報ごとに可変長の区画に区分された音声情報

列が記録された第1領域と、該音声情報列の各区画ごとに、該各区画の前半部分と後半部分との境界を示す境界位置識別情報が記録された第2領域とを、少なくとも備えている。

【0011】なお、当該音声情報記録媒体は、複数種類の音声情報列が記録された複数の領域を備えてもよい。すなわち、音声再生中の遅延感知機能は、上述の日本国特許第2581700号に示されたように複数種類の音声情報の切替え再生可能な場合、この発明に係る音声記録媒体（音声記録媒体に係る第2実施態様）は、所定の音声再生手段で再生出力されるべき複数の単語列から構成された1又は2以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された音声情報ごとに可変長の区画に区分された第1音声情報列が記録された第1領域と、及び該第1音声情報列と等価でかつ遅延した発音で構成された音声情報列であって、該第1音声情報列の各区画の音声情報と等価な音声情報ごとに可変長の区画に区分された第2音声情報列が記録された第2領域と、該第1及び第2音声情報列を該音声再生手段で切換えて再生すべく、切換え可能な各区画を当該記録媒体における該各区画の記録位置で示す記録位置識別情報が記録された第3領域と、そして、該1及び第2音声情報列における各区画ごとに、該各区画の前半部分と後半部分の境界を示す境界位置識別情報が記録された第4領域とを、少なくとも備えている。

【0012】また、この発明に係る音声記録媒体は、上述のいずれの構成であっても、区画に関する情報として、各区画中の発音状態に応じて（発音のまとまりごとに）区切られた前半部分と後半部分の境界を示す境界位置識別情報が、当該音声記録媒体の所定領域に記録されたことを特徴としている。

【0013】一方、この発明に係る音声再生方法は、音声再生中に発生する種々の再生方法を指示する割込み要求に対し、各区画にあらかじめ設定された上記境界位置識別情報を利用して、該割込み要求の対象を特定している。なお、割込み要求の再生指示には、リピート機能、上記音声記録媒体の各領域に格納された複数の音声情報を切替え再生する切替え再生機能、あるいはこれらの組み合わせ等が少なくとも含まれる。

【0014】上述のリピート機能を実現する場合、この発明に係る音声記録媒体の再生方法（再生方法に係る第1の実施態様）では、まず、所定の音声再生手段で再生出力されるべき複数の単語列から構成された1又は2以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された音声情報ごとに可変長の区画に区分された音声情報列が記録された第1領域と、該音声情報列の各区画ごとに、発音状態によって区切られた該区画の前半部分と後半部分の境界を示す境界位置識別情報が記録された第2領域とを、少なくとも備えた音声記録媒体が用意される。そして、記録された複数の区画のうち所

定区画の再生中に操作者からのリピート再生に関する割込み要求が発行されると、この割込み要求の発生タイミングが判断される。具体的には、再生中であった区画の先頭位置から境界位置までの該区画の前半部分で割込み要求が発生したか、あるいは該再生中であった区画の境界位置から次区画の先頭位置までの該区画の後半部分で割込み要求が発生したかが判断される。ここで、上記割込み要求が再生中であった区画の前半部分で発生したと判断されると、割込み要求が対象とする区画として再生中であった区画の前区画が特定される。一方、上記割込み要求が再生中であった区画の後半部分で発生したと判断されると、割込み要求が対象とする区画として再生中であった区画自身が特定される。

【0015】再生方法に係る第1実施態様では、以上のような割込み処理により特定された区画（操作者が意図した対象区画として特定された区画）の先頭から、音声再生を再開することによりリピート機能を実現する。

【0016】一方、当該記録媒体に記録された複数の音声情報列を切替え再生する場合、この発明に係る音声記録媒体の再生方法（再生方法に係る第2の実施態様）では、まず、所定の音声再生手段で再生出力されるべき複数の単語列から構成された1又は2以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された音声情報ごとに可変長の区画に区分された第1音声情報列が記録された第1領域と、該第1音声情報列と等価でかつ遅延した発音で構成された音声情報列であって、該第1音声情報列の各区画の音声情報と等価な音声情報ごとに可変長の区画に分割された第2音声情報列が記録された第2領域と、該第1及び第2音声情報列を該音声再生手段で切換えて再生すべく、切換え可能な各区画を当該記録媒体における該各区画の記録位置で示す記録位置識別情報が記録された第3領域と、そして、該第1及び第2音声情報列における各区画ごとに、発音状態によって区切られた該各区画の前半部分と後半部分の境界を示す境界位置識別情報が記録された第4領域とを、少なくとも備えた音声記録媒体が用意される。そして、これら区画のうち所定区画の再生中に第1及び第2音声情報列間での切換え再生に関する割込み要求が発生した場合、一旦割込み要求の発生タイミングが判断される。すなわち、この割込み要求が、再生中であった区画の先頭位置から境界位置までの該区画の前半部分で発生したか、あるいは該再生中であった区画の境界位置から次区画の先頭位置までの該区画の後半部分で発生したかが判断され、この判断結果に基づいて割込み対象となる区画が特定される。特定される区画は、この割込み要求が再生中であった区画の前半部分で発生したと判断された場合、該区画の前区画が特定され、割込み要求が再生中であった区画の後半部分で発生したと判断された場合には、該区画自身が特定される。

【0017】再生方法に係る第2の実施態様では、以上

のような割込み処理により操作者が意図した指示対象としての区画が特定されると、割込み要求が対象とする区画として特定された区画に対応する区画であって、第1及び第2音声情報列のうち上記割込み要求によって指示された音声情報列の区画の先頭から、音声再生が再開されることにより切替え再生機能を実現する。

【0018】なお、以上のように切替え再生可能な音声記録媒体の再生方法であっても、上述のリピート再生等の種々の機能をさらに備えることは可能である。

10 【0019】また、この発明に係る音声記録媒体は、ポータブルCDプレーヤーやCD-ROMドライバ等を備えたパーソナルコンピュータにより再生可能な情報記録媒体を含む。したがって、このような種々のデバイスにより再生されるべく、各音声情報の再生タイミングに対応して再生可能な画像情報、文字情報等を含んでもよい。

【0020】

【発明の実施の形態】以下、この発明の一実施例を図1～図18を用いて説明する。なお、図中同一部分には同一符号を付して説明を省略する。

20 【0021】（音声記録媒体に係る第1の実施態様）まず、この発明に係る音声記録媒体に記録されるべき音声情報列の基本的な構造を図1を用いて概念的に説明する。

【0022】この発明に係る音声記録媒体で記録される情報は、映画における出演者の会話、日常生活環境における会話等のように、長さの異なる複数のセンテンス（文）から構成され、また、各センテンス（各会話者の音声情報）の間に、音声再生されていない状況、雑音のみが再生されている状況、音楽（BGM）のみが再生されている状況等のランダムに発生する無音声期間が存在し得る一連の音声情報列である。したがって、この音声記録媒体に係る第1の実施態様（基本構成）は、所定の音声再生手段で再生出力されるべき複数の単語列から構成された1又は2以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された音声情報ごとに可変長の区画（以下、セグメントという）に区分された音声情報列が記録された第1領域を少なくとも備えている。

40 【0023】一般にネイティブスピーカーの英会話では、1センテンスは概ね3秒程度で発声されるため、記録されるべき音声情報列を構成するセグメントを決定する発音の節目を各センテンスの間に設定することで、図1(a)、(b)あるいは(d)に示されたように、音声情報列を構成する可変長セグメントをそれぞれ構成するのが妥当である。なお、会話中のセンテンスの中には図1(c)に示されたように、極端に短いセンテンスも含まれるが、このセンテンスも1つのセグメントを構成する。一方、図1(e)に示されたように、極端に長いセンテンスの場合には、接続詞や関係詞等の前が発音の

節目となるため、図 1 ( e ) に示されたようなセンテンスでは、連続する 2 つのセグメントで構成するのが妥当である。したがって、記録されるべき音声情報列のセグメントとは、発声上の区切り ( 息継ぎ位置 ) 又は言語上 ( 文法上 ) のなんらかの区切りにもとづいて分割された音声情報の記録単位であることを意味する。

【 0 0 2 4 】また、当該音声記録媒体は、上述のような特殊な音声情報の再生中に発生する可能性のある割込み要求を適切に処理すべく、上述のように発音の節目で分割された各セグメントの前半部分と後半部分の境界を示す境界位置識別情報が記録された第 2 領域をさらに備えている。具体的にこの境界位置識別情報は、例えば当該音声記録媒体のディレクトリ領域や各セグメントのヘッダ部に記録される。

【 0 0 2 5 】分割された各セグメントごとに設定される境界は、図 1 ( a ) 、 ( b ) あるいは ( d ) に示されたように、何等かの発音のまとまりがある最初の区切り S 1 、 S 2 、 S 4 に設定される ( すなわち、 S 1 、 S 2 、 S 3 、 S 4 は各区画ごとに発音状態に応じて設けられ、これにより各区画は前半部分と後半部分に区切られる ) 。なお、図 1 ( c ) に示されたように、センテンスが極端に短い場合には、境界位置 S 3 を分割されたセグメントの終了位置と一致するよう設定してもよい。また、図 1 ( e ) に示されたように、1 つのセンテンスが複数のセグメントに分割された場合も、各セグメントについて発音のまとまりと判断された前半部分と後半部分とに区切る境界を示す境界位置 S 5 、 S 6 をそれぞれ設定する。

【 0 0 2 6 】特に、この第 1 の実施態様に係る音声記録媒体の音声再生は、記録されたセグメントごとに順番に行われる。しかしながら、係る音声再生中にリピート指示等の操作者からの割込み要求があった場合には、一旦この割込み要求の発生タイミングが判断され、この判断結果に基づいて該割込み要求の対象となるセグメントが特定される。具体的には、この割込み要求が、再生中であったセグメントの先頭位置から上述のように設定された境界位置までの前半部分で発生したか、あるいは該再生中であったセグメントの境界位置から次に再生されるべき次セグメントの先頭位置までの後半部分で発生したかが判断される。ここで、上記割込み要求が再生中であったセグメントの前半部分で発生したと判断されると、割込み要求が対象とするセグメントとして再生中であったセグメントの前のセグメント ( 既に音声再生が終了したセグメント ) が特定される。一方、上記割込み要求が再生中であったセグメント後半部分で発生したと判断されると、割込み要求が対象とするセグメントとして再生中であったセグメント自身が特定される。

【 0 0 2 7 】この再生方法は、このように特定されたセグメント ( 操作者の指示対象として特定されたセグメント ) の先頭から、音声再生を再開することによりリピー

ト機能を実現する。

【 0 0 2 8 】 ( 音声記録媒体に係る第 2 の実施態様 ) 次に、この発明に係る音声記録媒体の第 2 の実施態様は、少なくとも 2 種類の音声情報列が記録されていることを特徴としている。すなわち、第 1 音声情報列は例えばネイティブスピーカが自然な速さで話す英語の音声情報からなり、この音声情報列は上述されたように発音の節目 ( センテンスの終りやセンテンス中の一息つける、発生上あるいは文法上の区切り ) で複数の可変長セグメントに分割されている。第 2 音声情報列は上記第 1 音声情報列の内容と等価な意味内容であるが別の音声情報であり、例えば単語を区切って話すゆっくりとした速さの英語の音声情報である。なお、この第 2 音声情報列も、複数の可変長セグメントから構成されている。

【 0 0 2 9 】ここで重要なことは、上記第 1 及び第 2 音声情報列はそれぞれ複数の可変長セグメントに区分されているが、互いにセグメントごとにその意味内容が対応していることである。例えば、第 1 音声情報列の 6 番目 ( 図 1 ( a ) では 6 2 1 番目 ) のセグメントがネイティブスピーカの話す " It's not much of a problem. " であるときは、第 2 音声情報列の 6 番目 のセグメントは各単語を区切って話す " It is not much of a problem. " となる。ただし、第 2 音声情報列と対応した内容でかつ別の音声情報からなるということは、言語上は同一の意味で発声の異なるものであることを示している。

【 0 0 3 0 】さらに、この第 2 の実施態様に係る音声記録媒体は、所定領域に、記録位置識別情報が記録されていることを特徴としている。この記録位置識別情報は、少なくとも、第 1 及び第 2 音声情報列における各セグメントが、当該音声記録媒体のどの位置に記録されているかを示す情報である。したがって、例えば第 1 音声情報列の 6 番目 のセグメント " It's not much of a problem. " に対応する第 2 音声情報列のセグメント " It is not much of a problem. " が、媒体のどの位置に記録されているかということは、この記録位置識別情報により認識することができる。

【 0 0 3 1 】その結果、第 1 及び第 2 音声情報列と記録位置識別情報は互いに無関係に記録されるのではなく、一定の関係をもって記録され、各音声情報列はセグメントを単位として有機的に組み合わせられている。すなわち、第 1 及び第 2 音声情報列は互いに対をなしており、これらをセグメントごとに関連させているのが記録位置識別情報である。なお、この実施態様では、記録位置識別情報は当該音声記録媒体のディレクトリ領域に記録されており、少なくとも各セグメントの先頭位置に関する情報を含んでいる。また、この実施態様では境界位置識別情報は、各セグメントのヘッダ部分に記録される。

【 0 0 3 2 】以上のような構造を備えた音声記録媒体 ( 第 2 の実施態様 ) の再生方法でも、記録されたセグメ

ントごとに順番に音声再生が行われるが、特に、この再生方法では、当該音声記録媒体に記録された第1音声情報列から第2音声情報列への再生切換え（あるいは第2音声情報列から第1音声情報列への再生切換え）が可能であることを特徴としている。なお、この再生切換え動作は、セグメントを単位として行われる。例えば、第1音声情報列の*m*番目のセグメントが再生されているときに第2音声情報列の再生指示が入力されると（割込み要求の発生）、記録位置識別情報に基づいて第2音声情報列の対応する*m*番目のセグメントを読み出し、その対応するセグメントの音声再生が実行される。また逆に、第2音声情報列から第1音声情報列への再生切換えも、上述した再生切換え動作と同様に各セグメント単位で行われる。

【0033】なお、上述の切換え動作において、切換え対象となるセグメントの特定は、割込みタイミングを上述された第1の実施態様に係る音声記録媒体の再生方法で説明されたように判断することにより行われる。また、この再生方法では、上述の再生切換え動作の他、リピート再生等の種々の変形が可能である。その代表的なものとして、いわゆる戻し指令がある。すなわち、再生中の停止命令により一時再生を中断した後に戻し指令が入力されたときは、指令された量だけ音声情報の読み出し位置を戻すことによりより操作者の希望に合った音声情報の再生が行われる。

【0034】（音声記録媒体の第3の実施態様）さらに、この発明に係る音声記録媒体の第3の実施態様について説明する。この第3の実施態様に係る音声記録媒体は、第1及び第2音声情報列の他、さらに第3音声情報列が当該音声記録媒体に記録されている点が、上記第2の実施態様に係る音声記録媒体と異なる。

【0035】ここで重要なことは、上記第3音声情報列は第1及び第2音声情報列の1又は2以上の可変長セグメントをひとまとまりとしたセグメント群に区分されていることである。換言すれば、この第3音声情報列の1つのセグメント群は第1及び第2音声情報列の1又は2以上のセグメントを包含しており、したがって、第3音声情報列の1つのセグメント群は第1及び第2音声情報列の1又は2以上のセグメントと対になっている。特に、この構成は図1（e）に示されたように、1つのセンテンスが複数のセグメントに区分された場合を想定している。

【0036】また、この第3の実施態様の音声記録媒体において、所定の領域に記録された記録位置識別情報には、上記第3音声情報列の内容の記録位置をもセグメント群ごとに示す情報も含まれている。したがって、第1、第2及び第3音声情報列と記録位置識別情報は互いに一定の関係をもって媒体に記録され、各音声情報列はセグメントあるいはセグメント群を単位として有機的に組み合わされている。なお、この第3の実施態様にお

ても、記録位置識別情報は当該音声記録媒体のディレクトリ領域に記録され、各音声情報列におけるセグメントの先頭位置に関する情報を含んでいる。また、この実施態様においても、境界位置識別情報は各セグメントのヘッダ部に記録される。

【0037】以上のような構造を備えた音声記録媒体（第3の実施態様）の再生方法は、基本的に上述された第2の実施態様の場合と同じであるが、第1及び第2音声情報列間での再生切換えの他、該第1及び第2音声情報列と第3音声情報列との間においても再生切換え動作を行う点が、第2の実施態様の場合と異なる。

【0038】例えば、第1音声情報列の再生中にネイティブスピーカの”It's not much of a problem.” が聴き取れなかったときは、再生中の第1音声情報列から第2音声情報列に再生を切換えることにより、ゆっくりと単語を区切って話す”It is not much of a problem.” を聴くことができる。そして、この日本語の意味や文法を知りたいときは、さらに、再生中の音声情報列から第3音声情報列へ再生を切換えればよい。もちろん、この再生方法においても、上記第2の実施態様に係る音声記録媒体の再生方法で説明された戻し指令や停止命令を組み合わせて使えるよう応用できることは言うまでもない。また、この再生方法においても、切換え再生及びリピート再生が可能であり、これら割込み要求が対象とするセグメントの特定は、上述された第1の実施態様の場合と同様に行われる。

【0039】（音声記録媒体の第4の実施態様）この発明に係る音声記録媒体の第4の実施態様は、基本的に上記第3の実施態様の場合と同様であるが、第1及び第2音声情報列の他、文字情報列が記録されている点が主に異なる。この文字情報列は、第1又は第2音声情報列に対応する内容の文字情報に相当しており、例えばネイティブスピーカが話す英語（音声）に対応する文字情報に相当している。

【0040】この文字情報列も、第1及び第2音声情報列の各セグメントと対応するセグメントに区分されている。また、この第4の実施態様に係る音声記録媒体においても、記録位置識別情報には、この文字情報列の記録位置を各音声情報列のそれぞれのセグメントごとにそれらの先頭位置に関する情報が含まれ、当該音声記録媒体のディレクトリ領域に記録される。また、各セグメントのヘッダ部には境界位置識別情報が記録される。したがって、第1及び第2音声情報列と文字情報列はそれぞれセグメント単位で対応することになる。なお、この第4の実施態様に係る音声記録媒体において、上述の第3の実施態様における第3音声情報列を記録情報として加えるときは、第1及び第2音声情報列と文字情報列の1又は2以上のセグメントは第3音声情報列の1つのセグメント群にも対応することになる。さらに、この第4の実施態様においても、上記記録位置識別情報には、各セグ

メントの先頭位置が含まれ、かつ当該音声記録媒体のディレクトリ領域に記録される。また、境界位置識別情報は各セグメントのヘッダ部に記録される。

【0041】以上のような構造を備えた音声記録媒体（第4の実施態様）の再生方法も、基本的に上記第2の実施態様の場合と同様であるが、第1又は第2音声情報列の再生中に文字情報列がディスプレイ表示される点が異なる。

【0042】例えば、第1音声情報列のセグメント”It's not much of a problem.”が再生されているときは、所定の表示部に”It's not much of a problem.”もしくは”It is not much of a problem.”がディスプレイ表示される。なお、この表示については再生中の音声情報列と時間的に完全に同期している必要はなく、文字が少しずつ遅れて表示されたり、あるいは少しずつ先に表示されたりしてもよい。また、この再生方法においても、切換え再生及びリピート再生が可能であり、これら割込み要求が対象とするセグメントの特定は、上述された第1の実施態様に係る音声記録媒体の再生方法と同様に行われる。

【0043】次に、この発明に係る音声記録媒体の具体的な構造を、図2～図5を用いて、以下詳細に説明する。

【0044】図2は、この本発明に係る音声記録媒体の例として、上述の第3の実施態様を英会話独習用に適用したときの各音声情報列A、B、Cと、その記録内容を説明するための図である。この図において、音声情報列Aはネイティブスピーカの話す英語の情報列（第1音声情報列）であり、複数のセグメント621、622から構成されている。音声情報列Bはゆっくりと単語を区切って話す英語の情報列（第2音声情報列）であり、これも先の音声情報列Aのセグメント621、622に対応する英単語あるいは句からなる複数のセグメントから構成されている。また、音声情報列Cは日本語の解説をする情報列（第3音声情報列）であり、この音声情報列Cに含まれるセグメント群は、各音声情報列A、Bの各セグメント621、622にそれぞれ対応している。

【0045】また、図3は、図2に示された態様における1セグメント当りの時間と容量の関係を説明するための表である。この表において、1秒間は6キロバイトの容量に対応している。例えば音声情報列Aのセグメント621では、”It's”の発声時間が0.2秒、その容量が1.2KB（キロバイト）、”not”の発声時間が0.1秒、その容量が0.6KB（キロバイト）、”much of a”の発声時間が0.4秒、その容量が2.4KB（キロバイト）、そして”problem”の発声時間が0.3秒、その容量が1.8KB（キロバイト）であり、セグメント621全体の発声時間は2.0秒、その容量は12KB（キロバイト）となる。なお、表中におけるカッコの数字は、境界位置を再生時間で示した値で

ある。すなわち、この音声情報列Aの場合は、図1(a)に示されたように、”not”の後に境界位置が設定されるため、セグメント621の再生開始から0.3秒後の位置に音声情報列Aを前半部分と後半部分に分ける境界が設定されていることを示している。また、他の音声情報列B、Cについても同様に示されているが、この図3の実施態様では、音声情報列Cには境界位置識別情報は設定されていない。

【0046】さらに、図4は、図2及び図3に示された態様におけるディレクトリ領域の記録内容を説明するための表である。この表において、ディレクトリ領域は、1セグメント当たり $9 \times 3 = 27$ バイト（B）で構成される。音声情報列A、B、Cはそれぞれ図2の音声情報列A、B、Cに対応している。また、1バイトのCは属性を示し、C=0は音声情報列A、C=64は音声情報列Bであることを意味する。また、C=128、129は音声情報列Cであることを意味し、特にC=129のとき、すなわちビット表現（8ビット（bit））で”10000001”のときは前のセグメントと同じ解説対象であることを示す（音声情報列Cの解説対象となる同じセグメント群に属していることを示し、例えば図1(e)のセグメント801、802の場合が相当する）。

【0047】位置情報のM、S、B（各1バイト）は産業界で標準になっているCD-ROM上の位置を表わすパラメータである。すなわちMは分、Sは秒、Bはブロックをそれぞれ示す。また、1ブロックは2,048バイトであり、75ブロックで1秒分を構成している。したがって、最大のは数はM=59、S=59、B=74となる。次の2バイトのSBはスタートバイトを示し、その次の3バイトのLLLは各セグメント全体の長さを示している。なお、位置を示すパラメータに分、秒を使う理由はCD-ROMはもともと音楽用として開発されたためであり、始めからの時間として記録位置を表現するようになっている。そのためCD-ROMを当該音声記録媒体として採用した場合には、この分と秒は再生時の時間とは全く無関係であり、単に記録媒体上の記録位置を表わしている情報にすぎないことになる。

【0048】その結果、例えば音声情報列Aにおけるセグメント621の”It's notmuch of a problem.”は、0分11秒3ブロックの826バイト目から6,000バイトの長さでネイティブスピーカの話す英語の音声情報が記録され、音声情報列Bにおける対応するセグメントは0分11秒3ブロックの2,026バイト目から17,400バイトの長さでゆっくりと単語を区切って話す英語が記録され、音声情報列Cのセグメント群は0分11秒6ブロックの1,282バイト目から72,000バイトの長さで日本語解説が記録される。なお、621、622等のセグメントナンバーはメモリ上にはなく、そのアドレスに対応している。また、各セグメント



の関係を示す記録位置識別情報は、このディレクトリ領域に含まれる。

【0049】さらに具体的には、第4図に示されたディレクトリ領域の記録内容から、当該音声記録媒体の0分11秒3ブロックにおける826バイト目から826+6,000-1=6,825バイト目までの領域には、セグメントが621で属性Cが0の音声情報列すなわちネイティブスピーカが話す”It's not much of a problem.”に相当する情報が記録される。また、当該音声記録媒体の0分11秒3ブロックにおける2,026バイト目から2,026+17,400-1=19,425バイト目までの領域には、セグメントが621で属性Cが64の音声情報列すなわち単語を区切ってゆっくり話す”It is not much of a problem.”に相当する情報が記録される。さらに、当該音声記録媒体の0分11秒6ブロックにおける1,282バイト目から1,282+72,000-1=73,281バイト目までの領域には、セグメントが621で属性Cが128の音声情報列すなわち日本語の解説に相当する情報が記録される。

【0050】このように、図4に示されたディレクトリ領域を設ければ、図3に示されたような再生時間及び容量で図2に示された各音声情報列が記録可能である。

【0051】次に、各セグメント621、622の前半部分と後半部分の境界を示す境界位置識別情報は、例えば図5(a)に示された可変長セグメントのヘッダ部に記録される。このヘッダ部は、図5(b)に示されたように、先頭から文字情報や画像情報の有無等を示すための1バイト領域(1B)、音声情報列A用に用意された領域であって情報列タイプ(音声情報列A、B等を区別するための情報)を示す1バイトデータ、そのデータ長を示す3バイトデータ、及び境界位置を示す1バイトデータから構成された5バイト領域(5B)、音声情報列B用に用意された領域であって情報列タイプを示す1バイトデータ、そのデータ長を示す3バイトデータ、及び境界位置を示す1バイトデータから構成された5バイト領域(5B)、音声情報列C用に用意された領域であって情報列タイプを示す1バイトデータ及びそのデータ長を示す3バイトデータから構成された4バイト領域(4B)、文字情報列D用に用意された領域であって情報列タイプを示す1バイトデータ及びそのデータ長を示す3バイトデータから構成された4バイト領域(4B)、同様に文字情報列D用に用意された領域であってアドレスを示す3バイトデータ及びそのデータ長を示す3バイトデータから構成された6バイト領域(6B)、他の情報列(タイプE)用に用意された4バイト領域(4B)、及び予備の3バイト領域(3B)からなる、32バイトの領域である。

【0052】なお、この明細書中で説明される実施態様では、他の情報列(タイプE)は使用していない。また、この実施態様では、境界位置識別情報は音声情報列

A、Bにのみ設定されている。境界位置を示す1バイト領域には、圧縮後の情報においてセグメント先頭からのバイト数を256で割った商がセットされる。例えばAトラック(音声情報列Aに相当する仮想トラック)の音声サンプリングレートを128キロビット/秒(16キロバイト/秒)とすると、この1バイト領域で先頭から4秒までの範囲の境界位置が表せ、その分解能は0.016秒となる。また、音声サンプリングレートが32キロビット/秒(4キロバイト/秒)の場合には、この1バイト領域で先頭から16秒までの範囲の境界位置が表せ、かつその分解能は0.064秒となり、実用上十分な性能が得られる。

【0053】次に、図6～図10を用いて、この発明に係る音声記録媒体の基本的な再生方法を説明する。

【0054】まず、図6は、この発明に係る音声記録媒体の再生方法実現するための再生装置の全体構成を示す斜視図である。この図からも分かるように、当該音声記録媒体は、例えばポータブルなCDプレイヤー(再生装置本体2)により再生可能なCD-ROMであり、この再生装置本体2はコード接続されたハンドセット8によりリモート制御される。このハンドセット8には少なくとも再生中のセグメント番号を表示する液晶ディスプレイ(LCD)等の表示部21や、各種制御用ボタン群240が設けられている。また、操作者は再生装置本体2で再生された音声情報をイヤホン13を介して聴くことができる。

【0055】また、図7は、図6に示された再生装置の構成を示すブロック図である。この図に示されたように、当該音声記録媒体1であるCD-ROMは再生機構20にセットされる。再生機構20はディスクインターフェイス(I/F)3及びバス4を介してCPU5に接続されている。また、バス4にはプログラムを格納するための例えば32キロバイト(KB)のROM6と、ディレクトリや音声情報列を一時的に格納するための例えば256キロバイトのRAM7とが接続されている。さらに、バス4には手動操作のためのハンドセット8との間で情報の授受を行なうハンドセットインターフェイス(I/F)9と、音声出力用のアンプ(AMP)10を介して外部端子11及びハンドセット8に接続されたD/Aコンバータ12に接続されている。なお、ハンドセット8には上述されたようにイヤホン13が接続されている。

【0056】図8(a)、(b)は、それぞれROM6及びRAM7のメモリ割り当て状況を説明するための図である。図8(a)に示されたように、32キロバイトのROM6にはプログラムが格納される。一方、図8(b)に示されたように、RAM7には、(50+50)=100キロバイトのバッファ(50ブロック分に相当)と、(75+75)=150キロバイトのディレクトリと、6キロバイト分のシステムエリアが割り当て

られる。したがって、RAM7には常時50ブロック分の音声情報列が保持され、かつ150キロバイト÷27÷5、555セグメント分のディレクトリ（音声情報列Aの部分のみで約30分間に相当）が保持される。

【0057】なお、上述の具体例では当該音声記録媒体としてCD-ROMを用いているが、その代表的なものの容量は552メガバイト（MB）である。CD-ROMではアドレスを表わすのに分、秒、ブロックの単位を用いている。また、1ブロックは2,048バイト、75ブロックは1秒、60秒は1分であるため、該CD-ROMのアドレスの最大の値は59分59秒74ブロックである。逆に、このCD-ROMの容量は2,048×75×60×60=552.96メガバイトである。このうち、最初から2秒分はCD-ROMのフォーマットとしてユーザは使えないので、正確には最大容量とし552.6528MBとなる。さらに、最初から20秒に相当するところまでディレクトリ領域が割り当てられると、3メガバイトのディレクトリ容量をCD-ROMに確保することができる。

【0058】次に、容量に関する計算例を示す。

【0059】音声のサンプリング・レートはADPCM方式による16キロサンプル/秒とし、1サンプル当り3ビットとする。このように仮定すると音声サンプル量は48キロビット/秒（Kbit/s）すなわち6キロバイト/秒（KB/s）となり、転送レートもこれに合わせる必要がある。なお、16キロサンプル/秒であれば8キロヘルツ（KHz）までのf特性があり、子音まで十分に記録できる。このように仮定すると、1時間の録音に6キロバイト×3,600秒=21.6メガバイトの容量が必要となる。一般に、CD-ROM1枚にはエラー・コレクションを入れて552メガバイト記録できる。ディレクトリ領域を除くと549メガバイトを各音声情報毎に記録されるデータ領域として使用できる。したがって、549÷21.6=25時間24分の音声情報を記録することが可能となる。そこで、英会話用として用いる場合には、ネイティブスピーカによる自然な速度で発音されている物語を1時間とすると、それを単語1つ1つを区切ってゆっくり発音している部分はその4倍で約4時間とみることができる。そして、解説の部分は合計で15時間あったとしても全部で20時間である。

【0060】次に、1時間の会話部分をいくつかのセグメント（区画）に分解できるかを考えると、1秒間に平均4個の区切りとして1時間で14,400個である。1個のセグメントのディレクトリには36バイト必要なので全体で約518キロバイトであり、これは上記3メガバイトのディレクトリの収納場所に充分収納でき、かつ1時間の物語の全てのディレクトリを格納することができる。

【0061】次に、図9は、図6及び図7に示されたハンドセット8とイヤホン13の構成を具体的に示す平面

図である。ハンドセット8の前面にはセグメント番号を表示する表示部21（LCD）、現時点の指示内容をLEDを利用して示す表示部22、再生モード（リピート再生が実施されている場合等の再生状態を示す）を切換えるためモードスイッチ81、各音声情報列A、B、Cの再生指令をするための指令スイッチ23A、23B、23C、機能指示するためのADVボタン24A、再生の戻し指令をするためのREVボタン24B、リピート再生を指示するためのREPボタン25A、再生の停止を指令するためのSTOPボタン25B、及びイヤホン13から出力される音量を調節するためのボリュームスイッチ28が設けられている。また、イヤホン13はコード26を介してハンドセット8に接続され、ハンドセット8はコード27を介して再生装置本体2に接続されている。

【0062】次に、この発明に係る音声記録媒体の再生方法における、基本的な再生シーケンスの一例を説明する。図10は、この基本的な再生シーケンスを説明するための図である。音声情報列Aの部分すなわちネイティブスピーカの英語を続けて聴くときは、この図10（a）に示されたように順次記録されたセグメントの音声情報が再生される。この場合は、図9に示されたボタン23Aを押してそのまま聴いているだけでよい。

【0063】一方、図10（b）に示された再生シーケンスでは、まず、セグメント621から順に音声情報列Aの再生が行われる。ここで、例えばセグメント623がよく聴き取れなかった時には、操作者はすぐにSTOPボタン25Bを押すことにより一旦音声再生を中断する。このとき、セグメント表示部21のセグメント番号は624になっている。そこで、REVボタン24Bを1回だけ押す。なお、REVボタン24Bは押しつづけていると連続してもどり、1回押すごとに表示部21に表示されたセグメント番号は1つずつ戻る。セグメントが624の状態ではREVボタン24Bが1回押されると、表示部21に表示されたセグメント番号は623になる。

【0064】続いて、スイッチ23Bが押されると（切換え再生指示）、音声情報列Aのセグメント623に対応した音声情報列Bのセグメント623が再生され、1つ1つの単語を区切ってゆっくり発音している音声聴える。そのままにしておくと続けて音声情報列Bのままセグメント624の再生動作が進行する。さらに、この音声情報列Bのセグメント624を聴いたところでSTOPボタン25Bが押された後、再度REVボタン24Bが押されると表示部21のセグメント番号は623になる。この状態でスイッチ23Cが押されると、音声情報列Cの日本語の解説が再生される。解説の部分はいくつかの単語がまとまった句について解説していることもあるので、いくつかのセグメント番号で一つの解説ということもある属性Cを示すバイトの最下位ビットが1の

もの（属性Cが129のもの）はそれを示し、同一の音声情報列の前のセグメント番号と同じ解説対象であることを意味している。ここに当たったら読みとばすこともできる。

【0065】次に、図11～図13を用いて、この発明に係る音声記録媒体の他の実施態様を説明する。

【0066】図11は、この発明に係る音声記録媒体におけるディレクトリ領域の他の構成例を示す図である。この図に示されたディレクトリは、1セグメント当たり15バイトで構成されている。すなわち、1バイトの属性Cと、2バイトのセグメントナンバーSSと、1バイトのサフィックスナンバーNと、1バイトの分Mと、1バイトの秒Sと、1バイトのブロックBと、そして3バイトのセグメント長LLLの計12バイトである。

【0067】また、属性Cの8ビットにおいて、第1ビット（最上位ビット）が1のときはセグメントのスタートを意味し、0のときはその他の状態であることを意味する。また、この属性Cにおける第2、3ビットで表現される値が0のときは音声情報列Aであることを意味し、1のときは音声情報列Bであることを意味し、2のときは音声情報列Cであることを意味する。さらに、第4ビットが0のときはサフィックスが無いことを意味し、1のときはサフィックスが有ることを意味する。第5ビットが1のときは前のセグメントと同じセグメント群に関連づけられていることを意味し、0のときは同じセグメント群には関連づけられていないことを意味する。

【0068】図12は、図11に示されたディレクトリに対応した各音声情報列A、B、Cのセグメントの分割状況を説明するための図である。そして、これが図2に示された例と異なる点は、サフィックスを付したセグメントで区分されている点である。

【0069】また、図13は、以上説明されたように、8ビットの属性Cの第2、3ビットの値と分割された各区画に相当するセグメントとの関係を示す図である。

【0070】さらに、図14及び図15を用いて、この発明に係る音声記録媒体の他の実施態様の説明しておく。なお、この他の実施態様が先に説明した図11～図13の実施態様と異なる点は、この実施態様が音声情報だけでなく文字情報の記録、再生をも行なうようにしている点である。

【0071】第14図はそのディレクトリの構造を説明するための図である。この図において、先に説明した図11のディレクトリと異なる点は、属性Cの第2、3のビットに文字情報列Dに関する情報が付加される点である。このビットが3であるときは、文字情報列Dが当該音声記録媒体に所定のコードで記録されていることを意味する。

【0072】図16は、この実施態様に係る音声記録媒体の再生装置に用いられるハンドセットの構成を示す平

面図である。なお、この図15のハンドセットと図9のハンドセットとの差異は、文字情報列Dについても表示可能な態様である点である。具体的に、ハンドセット8には、文字表示を指示するための操作ボタン23Dとともに、LCDで構成される文字表示部210が設けられている点である。また、この実施態様では、文字の表示をオン、オフさせるための文字表示ボタンもさらに設けられている。このようなハンドセット8では、文字表示部210で文字情報列Dを表示するか否かの設定が可能である。

【0073】次に、図11～図13に示された他の実施態様の作用の特色を説明する。

【0074】この実施態様では、音声情報列Aの再生中は情報列Aの発音のスピード、すなわちLLLの長さに応じてその文字情報列Dの現われるスピードも制御される。つまり、このセグメントの発音の始まりから終わりまでの間に、文字が表示部210に現われ始めて完全に出し終るようにする（発音に完全に同期させて文字を表示させる）。

【0075】次に、音声情報列Bの再生中は音声情報列Aの部分の再生に比べて数倍の時間をかけて発音されているので、この時の音声情報列Bの発音の長さに同期させて文字情報列Dを表示させる。実際には、文字の表示の方を早めに出す（音声出力の方を少し遅らせて出す）方が使う人にとって便利なことも考えられるため、この点は任意に設定すればよい。

【0076】次に、図14及び図15に示された実施態様として、情報記録媒体をランダムアクセスのできない磁気テープとした場合を説明する。

【0077】この実施態様においても、基本的な情報の記録、再生の態様は上述の実施態様と同様であるが、磁気テープではデータの飛び越し再生が容易でないため、実用的な再生装置とするためには、情報を一時的に記憶しておくためのバッファを設けるのが望ましい。具体的には、前述の仮想トラックの音声情報列A、B、C（3トラック）に相当する各小単位を、 $A_1$ 、 $B_1$ 、 $C_1$ 、 $A_2$ 、 $B_2$ 、 $C_2$ 、 $A_3$ 、 $B_3$ 、 $C_3$ 、…というように、対応する音声情報列A、B、Cの各セグメントを1つのかたまりとして順番にならべる。そして、音声情報列Aを再生する時は $A_1$ 、 $A_2$ 、 $A_3$ というように音声情報列Aのセグメントのみ取り出して再生し、音声情報列Bを再生する時は $B_1$ 、 $B_2$ 、 $B_3$ 、…というように音声情報列Bのセグメントのみ取り出して再生する。音声情報列Cについても同様である。この時、セグメント $A_1$ の次にセグメント $A_2$ を切れ目なく再生するためにはこれらセグメントの間に存在するセグメント $B_1$ 、 $C_1$ を飛びこさなければならない。これは、通常のテープ走行では容易でないため、磁気テープからは早めにバッファの中に情報を取ってきておく。そして、飛び越して再生するのはバッファの中で行うようにする。

【0078】この磁気テープを、いわゆるDAT（デジタル・オーディオ・テープ）としたときには、次のようになる。まず、音声情報列Aのセグメントとして、Aトラックに記録されるネイティブ・スピーカの発音をある単位で区切る。例えば、ここでは1秒毎の一定時間に区切ることとする。そして、トラックB（音声情報列B）としては、ゆっくり（例えば平均1/3の速度で）話している内容のものとする。Aトラックの1秒に相当するところをBトラック用に4秒分とれるよう領域を確保する。そして、解説用のCトラックとしては9.4秒分とれるよう領域を確保する。なお、上記の秒数は一例であり、これに限定されるものではない。つまり、14.4秒を1つの単位とする。すなわち、14.4秒はA、B、Cトラック（各音声情報列A、B、Cに相当）の各対応する3個のセグメントをひとまとめたものである。ここで注意を必要とするは、Aトラック用の1秒分はオリジナルのものを1秒づつ切っているだけで、もとの1秒と完全に一致していることである。Bトラックの4秒は4秒以内であればよく、3秒で終わって1秒あまってもよい。なお、この余ったところ（1秒分）を再生する時に飛ばす方法については後述する。

【0079】ロータリ・ヘッド型DATの場合は次のようになる。すなわち、このロータリ・ヘッド型DATでは、ロータリ・ヘッドの1回転で2トラックを記録又は再生する方式が一般的である。そして、1トラックに2,880バイトの音声情報を記録できるようになっている。したがって、30トラックを1区切りとすると、86,400バイトとなる。この中に、上述の実施態様でも説明された音声のサンプリング・レート=48 (Kbit/s) で音声記録すると、 $86,400 \text{ (バイト)} \div 6,000 \text{ (バイト)} = 14.4 \text{ (s)}$  だけ記録できる。

【0080】次に、バッファについて考察する。

【0081】1MBのRAMを使用すると $10,480,576 \div 86,400 = 12,135$ となるので、12個のバッファを設けることができる。12個のバッファを機能的な意味でリング状にならべ、セグメント情報を常に2個づつ先取りしてテープからバッファに転送する。再生はこのバッファに転送された情報を利用して行われる。任意のところで止めた時には、10個のバッファが止めた時点より前の音声（既に再生音としてスピーカから出してしまったもの）が残っている。すなわち、Aトラック上の音声にして10秒の音声（テープを巻き戻すことなく、バッファから繰り返し再生することができる）が残っている。

【0082】このようにすることにより、10個のバッファに残っている10秒間の音声情報については、前述したCD-ROMを用いた実施例と同じく、任意のセグメントにおいてA、B、Cのどの仮想トラックにも容易

に移ることができる。

【0083】ただし、CD-ROMのようにランダム・アクセスができないので、A、B、Cの3つのトラックの対応するセグメントの時間数が、常に14.4秒以内にしなければならない。この時、仮想トラックAの部分は、常に一定の時間幅（例えば1秒毎）になるように、機械的に区切られていてもよい。例えば、Bトラックの部分は2秒の時もあるし5秒の時もあるので、可変になるようにする。14.4秒からA、Bトラックの時間を引いた時間がCトラックに割り当てられる時間である。なお、B、Cトラックのいずれも、区切りの中のAトラックに完全に対応した部分を記録する必要はなく、近くの複数の区切りを通して、あるまとまりができていればよい。

【0084】DATにおいてはサブコード領域があるので、各区切り（14.4秒/30トラック毎の）と、各区切りの中のA、B、Cトラックの境目を示すバイト数の情報は、このサブコード領域に記録することができる。

【0085】次に、DATに記録可能な音声情報量を具体的に計算する。

【0086】一般に、1時間用のDATには240,000トラックが記録できる。30トラックを1区切りに行っているが、次の6トラックを休みのトラックに使い、テープを止めている時にロータリ・ヘッドが多数回通っても、音声情報の記録されている部分にキズをつけないようにする。このようにするためには、36トラックが1単位となる。つまり、全体では6,666単位が記録でき、 $6,666 \times 14.4 \text{ 秒} = 126 \text{ 時間 } 39 \text{ 分 } 50 \text{ 秒}$ となる。このうち、Aトラック分として1時間に割り当てるので、B、Cトラックに25時間39分50秒を使えることになる。ゆっくり発音したBトラックが4時間分使ったとしても、説明用のCトラックに21時間以上割り当てることができ、十分な長さが確保できる。

【0087】なお、上記のA、B、Cトラックのセグメントのとり方は、前述のCD-ROMを応用したものと同じくAトラックは1秒等一定時間幅で区切るだけでなく、秒程度を目安とし、何らかの発音上の切り目があるところで切ってもよい。すなわち、記録する時の編集方針により変化してもよい。したがって、CD-ROMのようにランダム・アクセスのできるものでも、一定時間幅で区切ってセグメントを区分することも可能である。

【0088】次に、この発明に係る音声記録媒体の再生方法、特にリピート再生要求や切替え再生に関する割込み処理を図16～図18を用いて説明する。

【0089】図16は、この発明に係る音声記録媒体の再生方法における、割込み対象セグメントの決定動作を説明するためのタイムチャートである。

【0090】通常、この発明に係る音声記録媒体の再生方法では、あらかじめ指示された音声情報列のセグメン

ト (…、N-1、N、N+1、…) を順に再生していく。この再生途中で操作者 (例えば、英語学習者等) からリピート再生や切替え再生等に関する割込み要求 (図中、R1、R2、R3、…で示す) が発生したとき、まず、割込み要求の対象となったセグメントが特定される。すなわち、図16に示されたように、セグメントNが再生されているときに、係る割込み要求が先に説明された各セグメントのヘッダ部に記録された境界位置識別情報で示される境界位置  $S_i$  よりも前か後かが判断される。なお、図中において各セグメント内の境界位置は、それぞれ…、 $S_{i-1}$ 、 $S_i$ 、 $S_{i+1}$ 、…で示されている。さらに、この割込み要求の発生タイミングに関する判断結果に基づいて、当該割込み要求の対象セグメントとして再生中であったセグメントNかあるいはその前のセグメントN-1が特定される。発生した割込み要求の対象セグメントが特定されることにより、該割込み要求の指示内容にしたがって、この特定されたセグメントの先頭か、あるいは該特定されたセグメントに対応する他の音声情報列のセグメントの先頭から、音声情報の再生動作が再開される。

【0091】具体的に、例えばある音声情報列のセグメントNの再生中に所望の割込み要求R1が発生した場合、この割込み要求R1の発生タイミングはセグメントNの前半部分で発生したことになる。この場合は操作者の指示遅れと判断して再生中であったセグメントNの1つ前のセグメントN-1を割込み要求の対象セグメントとして特定する。一方、再生中であったセグメントNの後半部分で割込み要求R2が発生した場合は、この再生中であったセグメントN自身が割込み要求の対象セグメントとして特定される。なお、他のセグメント例えば図16に示されたセグメントN+1の再生中に発生した割込み要求R3の場合も同様に、境界位置  $S_{i+1}$  を基準に割込み要求の発生タイミングを判断し、割込み要求の対象セグメントが特定される。

【0092】図17は、この発明に係る音声記録媒体の再生方法における、リピート再生割込み処理 (再生方法に係る第1実施態様) を説明するためのフローチャートである。

【0093】この再生方法で用意される音声記録媒体には、少なくとも、所定の音声再生手段で再生出力されるべき複数の単語列から構成された1又は2以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された音声情報ごとに可変長のセグメントに区分された音声情報列と、該音声情報列における各区画ごとに、発音状態によって区切られた該各区画の前半部分と後半部分との境界を示す境界位置識別情報が、それぞれ所望の領域に記録されていればよい。

【0094】まず、以上のような音声記録媒体の再生中において、リピート再生に関する割込み要求が発生すると、一旦再生中であったセグメントの情報 (セグメント

ナンバー) が所定の保持手段によって保持されるとともに (ステップST1)、当該割込み要求の発生位置に関する情報も一旦保持される (ステップST2)。そして、再生中であったセグメントのヘッダ部に記録されていた境界位置と割込み発生位置とが時間軸に沿って比較される (ステップST3)。このとき、割込み要求が境界位置よりも後で発生した場合には、再生中であったセグメント自身が当該割込み要求の対象セグメントとして特定され、当該特定されたセグメントの先頭位置情報がディレクトリから読み出されて音声再生動作が再開される (ステップST4)。一方、割込み要求が境界位置よりも前で発生した場合には、再生中であったセグメントより1つ前のセグメントが当該割込み要求の対象セグメントとして特定され、当該特定されたセグメントの先頭位置情報がディレクトリから読み出されて音声再生動作が再開される (ステップST5)。

【0095】なお、この実施態様では、指示されたリピート再生動作は、次の新たな指示があるまで、特定されたセグメントの音声情報が繰り返し再生される。

【0096】次に、この発明に係る音声記録媒体の再生方法における、切替え再生割込み処理 (再生方法に係る第2の実施態様) を、図18のフローチャートを用いて説明する。

【0097】この実施態様で用意される音声記録媒体は、所定の音声再生手段で再生出力されるべき複数の単語列から構成された1又は2以上の文に対応する音声情報列であって、発音の節目でそれぞれ分割された音声情報ごとに可変長セグメントに区分された音声情報列Aと、この音声情報列Aと等価でかつ遅延した発音で構成された音声情報列であって、音声情報列Aの各セグメントの音声情報と等価な音声情報ごとに可変長セグメントに分割された音声情報列Bと、音声情報列A、Bを切替えて再生すべく、切換え可能な各セグメントを当該記録媒体における該各セグメントの記録位置で示す記録位置識別情報と、そして、音声情報列A、Bにおける各セグメントごとに、発音状態に応じて区切られた (発音のまとまりと判断された) 前半部分と後半部分との境界を示す境界位置識別情報とが、少なくとも記録されていればよい。

【0098】まず、以上のような音声記録媒体の再生中において、切替え再生に関する割込み要求が発生すると、一旦再生中であったセグメントの情報 (セグメントナンバー) が所定の保持手段によって保持されるとともに (ステップST6)、新たに指示された音声情報列の情報が保持される (ステップST7)。また、当該割込み要求の発生位置に関する情報も一旦保持される (ステップST8)。そして、再生中であったセグメントのヘッダ部に記録されていた境界位置と割込み発生位置とが時間軸に沿って比較される (ステップST9)。このとき、割込み要求が境界位置よりも後で発生した場合に

は、再生中であったセグメント自身が当該割込み要求の対象セグメントとして特定され、割込み要求で指示された音声情報列のセグメントであって、該特定されたセグメントの先頭位置情報がディレクトリから読み出されて音声再生動作が再開される（ステップST10）。一方、割込み要求が境界位置よりも前で発生した場合には、再生中であったセグメントより1つ前のセグメントが当該割込み要求の対象セグメントとして特定され、割込み要求で指示された音声情報列のセグメントであって、当該特定されたセグメントに対応するセグメントの先頭位置情報がディレクトリから読み出されて音声再生動作が再開される（ステップST11）。

【0099】なお、この発明は上述の各実施態様に限定されるものではなく、種々変形が可能である。

【0100】例えば、ビデオ・ディスクやビデオ・テープとの連動でもこの発明が適用できる。すなわち、普通のスピード（ネイティブ・スピーカの発声スピード）のAトラックに例えば映画のサウンド・トラック部を入れておき、Bトラックにゆっくりした発音を入れておき、Cトラックに日本語解説を入れておく。そして、Aトラックの再生を行ないながら、理解できなかったところで止め、少し戻し（この時、画面は固定のままでも実用に耐える）てBトラックを聴き、それでも分らなければCトラックを聴く。なお、Aトラックを再生する時だけ画面と同期させればよい。

【0101】また、パソコンなどにおいてこの発明を用いることもできる。すなわち、CAI（Computer Aided Instruction）との連動やCDI（Compact Disk Interactive）等高度のプログラマブルなものとも結合すれば、より高度の応用が可能である。また、パソコンの記憶装置を上記のCDやDATの代りに用いてこの発明を実施することも可能である。

【0102】さらに、英会話などに限らず、詩吟、法律用のものにも適用することができ、情報列は3種又は4種のものに限らず、それ以上であってもよい。

【0103】

【発明の効果】以上のようにこの発明によれば、各種音声情報とともに、該音声情報列を構成する可変長セグメントに関する情報として、記録媒体における該各セグメントの発音状態に応じて区切られた前半部分と後半部分との境界を示す境界位置識別情報が少なくとも記録された音声記録媒体を用意し、所定の割込み要求のタイミングが再生中であったセグメントの境界位置の前か後かを判断し、該判断結果に基づいて割込み要求の対象セグメントを特定するしているため、操作者の指示遅れも考慮されたフレキシブルな再生動作が実現できるという効果がある。

【図面の簡単な説明】

【図1】この発明に係る音声記録媒体に記録されるべき

音声情報列含む各種情報を概念的に説明するための図である。

【図2】この発明に係る、英会話独習用に適用された音声記録媒体（第3の態様）の各音声情報列と、その記録内容を説明するための図である。

【図3】図2に示された各音声情報列について、各セグメント当りの時間と容量との関係を説明するための表である。

【図4】図2及び図3に示された音声記録媒体におけるディレクトリ領域の記録内容（記録位置識別情報を含む）を説明するための表である。

【図5】この発明に係る音声記録媒体に記録されるべき可変長セグメントの構成（境界位置識別情報を含む）を示す図である。

【図6】この発明に係る音声記録媒体の再生方法を実現する再生装置の全体構成を示す斜視図である。

【図7】図6に示された再生装置の構成を示すブロック図である。

【図8】図7に示されたROM及びRAMのメモリ割り当て状況を説明するための図である。

【図9】図6に示されたハンドセット及びイヤホンの全体構成を示す平面図である。

【図10】この発明に係る音声記録媒体の再生方法における基本的な再生シーケンスを説明するための図である。

【図11】この発明に係る音声記録媒体におけるディレクトリ領域の他の構成例を示す図である。

【図12】図11に示されたディレクトリ領域の記録内容に対応した各音声情報列のセグメントの分割状況を説明するための図である。

【図13】図12に示された分割状況において、ディレクトリ領域の記録内容のうち8ビットの属性Cにおける第2及び第3ビットの値と分割された各セグメントとの関係を説明するための図である。

【図14】この発明に係る音声記録媒体におけるディレクトリ領域の他の構成例を示す図である。

【図15】図12に示されたディレクトリを有する音声記録媒体の再生を行う再生装置の、ハンドセットの概略構成を示す平面図である。

【図16】この発明に係る音声記録媒体の再生方法における、割り込み対象セグメントの決定動作を説明するための図（タイムチャート）である。

【図17】この発明に係る音声記録媒体の再生方法における、リピート再生に関する割り込み処理を説明するためのフローチャートである。

【図18】この発明に係る音声記録媒体の再生方法における、切換え再生に関する割り込み処理を説明するためのフローチャートである。

【図19】従来の音声記録媒体の再生方法における、リピート割り込み動作を説明するための図（タイムチャー

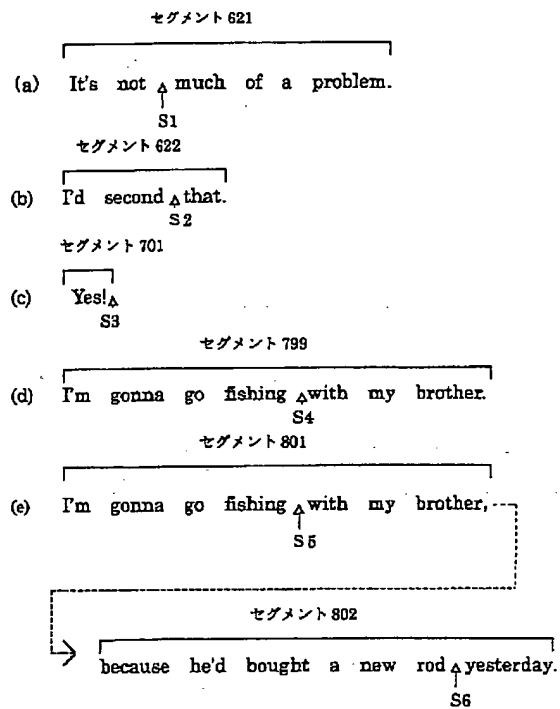
ト)である。

【符号の説明】

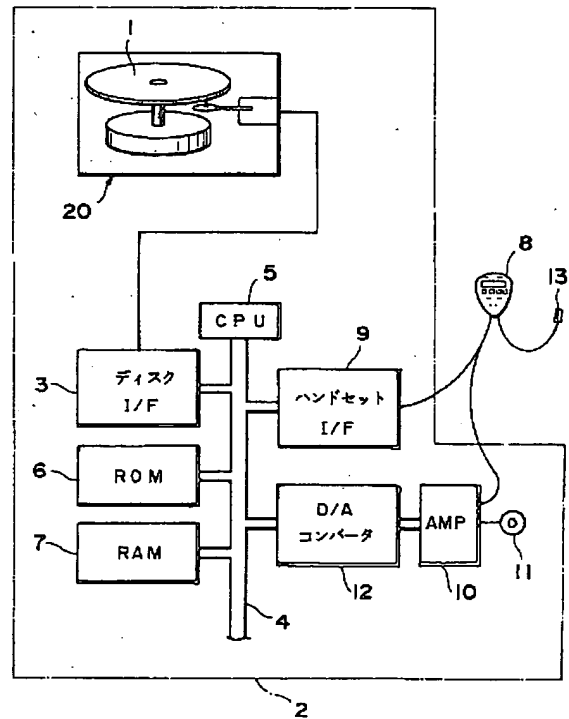
1…記録媒体、2…再生装置、8…ハンドセット、13

…イヤホン、21、210…表示部、23A、23B、  
23C、23D…再生情報列指示スイッチ、25A…リ  
ビート指示スイッチ。

【図1】



【図7】



【図2】

記録内容	
A	<div> <div>"It's not much of a problem."</div> <div>I'd second that."</div> <div>621</div> <div>622</div> </div>
B	<div> <div>"It is not much of a problem. I would second that."</div> <div>621</div> <div>622</div> </div>
C	<div> <div>(621)</div> <div>"It" とは前の〇〇を指しており、"not much a problem" は慣用句で『問題がない』とか『心配ない』というような意味です。</div> </div> <div> <div>(622)</div> <div>次に、"I'd" とは "I would" をつめた言い方であり、同じように "I could" も "I'd" とつめて言えます。</div> <div>"second" とは『支持する』、『賛成する』という意味があり、"I'd second that" で『同感です』という意味になります。</div> </div>

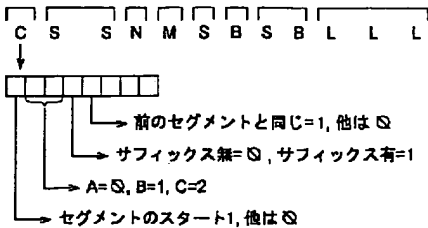
【図13】

2, 3ビット目	セグメント
0	621
1	621-1
1	621-2
2	621-1
2	621-2
2	621-3
2	621-4
0	622
1	622
2	622-1
2	622-2

【図 3】

セグメント		621	622
時間 (秒)	A	2.0 (0.3)	1.3 (0.4)
	B	3.9 (1.1)	2.6 (1.4)
	C	12	25
容量 (KB)	A	12	7.8
	B	23.4	15.6
	C	72	150

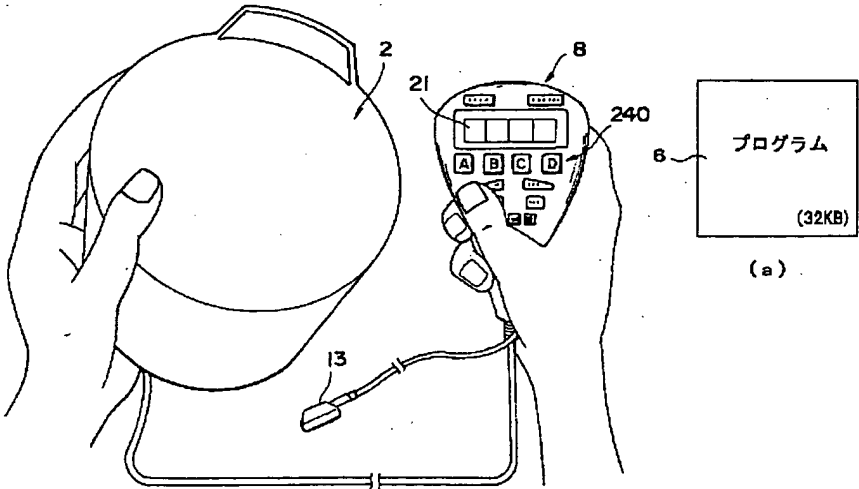
【図 11】



【図 4】

情報列	セグメント	カテゴリー	位置情報			スタートバイト	セグメント量 (Byte)
	SS	C	M	S	B	SB	LLL
A							
A	621	0	0	11	3	826	6,000
B		64	0	11	3	2,028	17,400
C		128	0	11	6	1,282	72,000
A	622	0	0	11	49	2,018	4,800
B		64	0	11	50	570	12,600
C		128	0	11	52	1,274	150,000

【図 6】



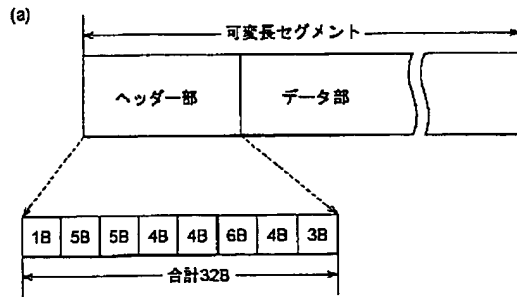
【図 8】

トラック用	(50KB)
ダブルバッファ	(50KB)
ディレクトリ用	(75KB)
ディレクトリ用	(75KB)
システムエリア	(8KB)

(b)



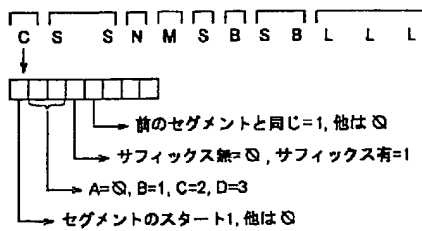
【図 5】



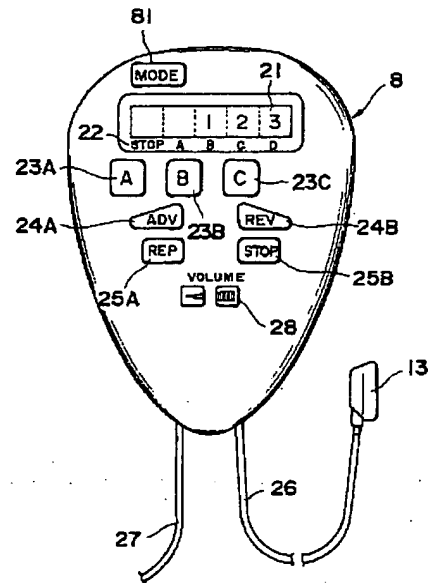
(b)

ヘッダー (合計32B)	情報列	内 容
1B		文字情報や画像情報の有無等 (1B)
5B	A	情報列タイプ (1B), データ長 (3B), 境界位置 (1B)
5B	B	情報列タイプ (1B), データ長 (3B), 境界位置 (1B)
4B	C	情報列タイプ (1B), データ長 (3B)
4B	(D)	情報列タイプ (1B), データ長 (3B)
6B	D	アドレス (3B), データ長 (3B)
4B	E	
3B	予備	

【図 14】

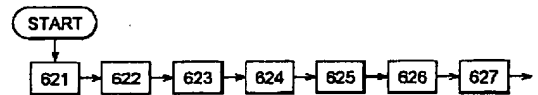


【図 9】

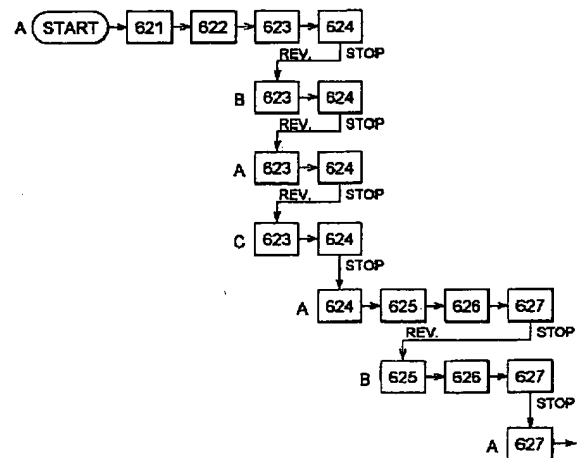


【図 10】

(a)



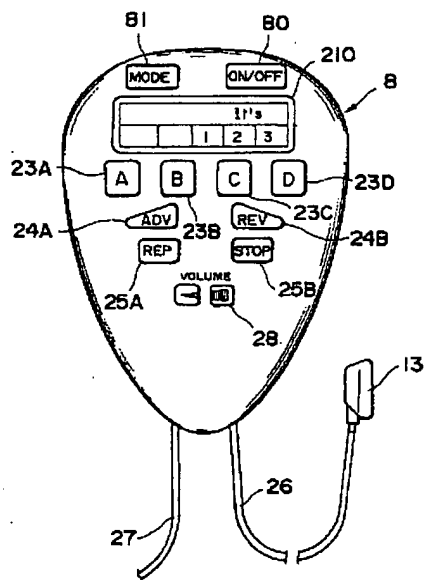
(b)



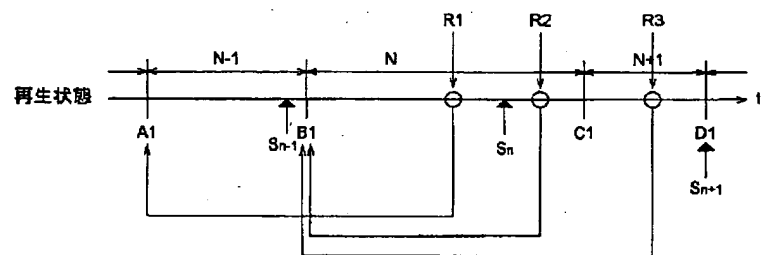
【図 12】

情報列	セグメント分割
A	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid black; padding: 2px; text-align: center;">It's not much of a problem.</div> <div style="border: 1px solid black; padding: 2px; text-align: center;">I'd second that.</div> </div> <div style="display: flex; justify-content: space-around; margin-top: 5px;"> <span>621</span> <span>622</span> </div>
B	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid black; padding: 2px; text-align: center;">It's not much of a problem.</div> <div style="border: 1px solid black; padding: 2px; text-align: center;">I'd second that.</div> </div> <div style="display: flex; justify-content: space-around; margin-top: 5px;"> <span>621-1</span> <span>621-2</span> <span>622</span> </div>
C	<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="border: 1px solid black; padding: 2px; text-align: center;">621-1</div> <div style="border: 1px solid black; padding: 2px; text-align: center;">621-2</div> <div style="border: 1px solid black; padding: 2px; text-align: center;">621-3</div> <div style="border: 1px solid black; padding: 2px; text-align: center;">621-4</div> <div style="border: 1px solid black; padding: 2px; text-align: center;">622-1</div> <div style="border: 1px solid black; padding: 2px; text-align: center;">622-2</div> </div>

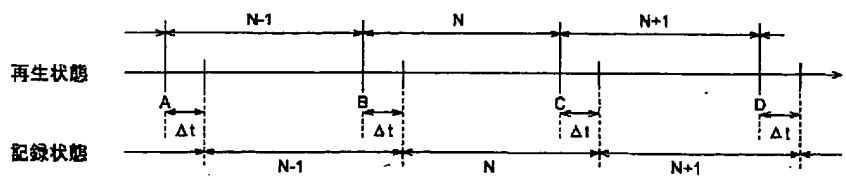
【図 15】



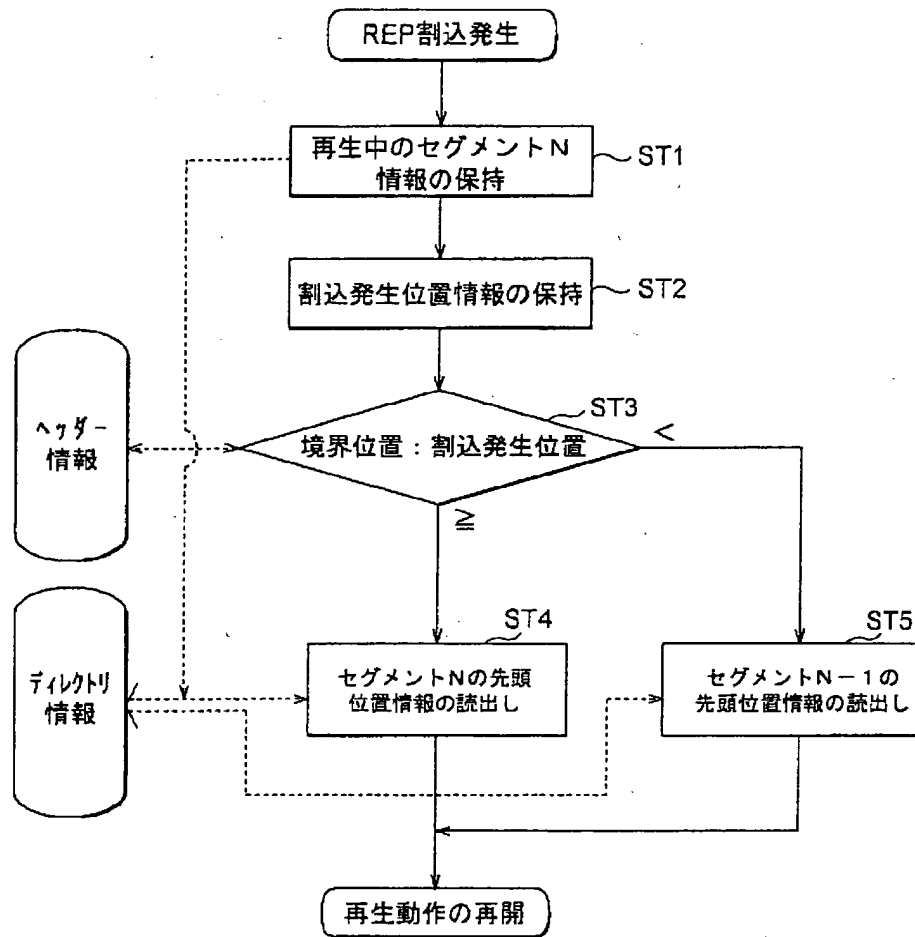
【図 16】



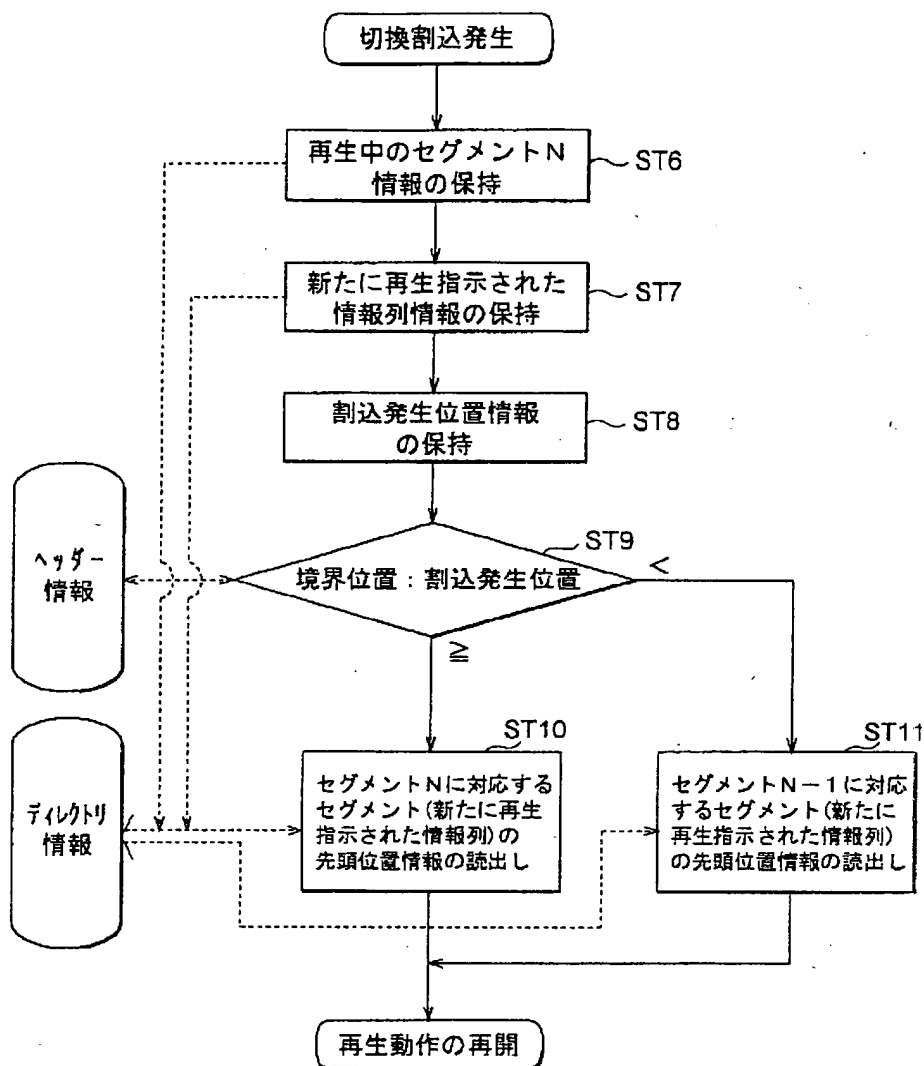
【図 19】



【図17】



【図18】



*Date: January 9, 2004*

*Declaration*

*I, Mariko Uchida, a translator of Fukuyama Sangyo Honyaku Center, Ltd., of 16-3, 2-chome, Nogami-cho, Fukuyama, Japan, do solemnly and sincerely declare that I understand well both the Japanese and English languages and that the attached document in English is a full and faithful translation, of the copy of Japanese Unexamined Patent No. Hei-11-134791 laid open on May 21, 1999.*

*Mariko Uchida*

*Mariko Uchida*

*Fukuyama Sangyo Honyaku Center, Ltd.*

SPEECH SOUND RECORDING MEDIUM, REPRODUCING METHOD THEREOF, AND  
INTERRUPT HANDLING METHOD DURING SPEECH SOUND REPRODUCTION  
Japanese Unexamined Patent No. Hei-11-134791

Laid-open on: May 21, 1999

Application No. Hei-9-294384

Filed on: October 27, 1997

Applicant: Kanars Data Corporation

Inventor: Hiroshi SEKIGUCHI

Patent Attorney: Yoshiki HASEGAWA et al.

#### SPECIFICATION

[TITLE OF THE INVENTION] Speech sound recording medium,  
reproducing method thereof, and interrupt handling method  
during speech sound reproduction

#### [ABSTRACT]

[Object] An object of the present invention is to provide a  
speech sound recording medium having a structure to enable  
flexible speech sound reproduction with an operator's  
instruction delay taken into consideration even when a series  
of speech sound data made up of a plurality of sentences  
different in length is reproduced, a reproducing method thereof,

and an interrupt handling method.

[Solution means] A speech sound data recording medium on which, together with speech sound data, as data concerning variable-length sections to make up the speech sound data sequence, at least boundary position identification data indicating boundary positions ( $S_{N-1}$ ,  $S_N$ ,  $S_{N+1}$ ) to split the respective sections ( $\dots$ ,  $N-1$ ,  $N$ ,  $N+1$ ,  $\dots$ ) of the recording medium into a first half part and a second half part judged to be cohesiveness of speech sound has been recorded is prepared, and whether timing of a predetermined interrupt request ( $R1$ ,  $R2$ ,  $R3$ ) is before or after the boundary position ( $S_{N-1}$ ,  $S_N$ ,  $S_{N+1}$ ) of a section during reproduction is judged, and based on a result of the judgement, a target section of the interrupt request is specified.

[WHAT IS CLAIMED IS;]

[Claim 1] An interrupt handling method during speech sound reproduction for executing handling of an interrupt request to instruct various reproducing methods generated during speech sound data reproduction of a speech sound recording medium, comprising: at least, a first region on which a speech sound data sequence, which is a speech sound data sequence corresponding to one or more sentences made up of a plurality of word sequences to be reproduced and outputted by a

predetermined speech sound reproduction means, sectioned into variable-length sections for each of speech sound data respectively divided at pauses in speech sound has been recorded; and a second region on which, for each section of the speech sound data sequence, boundary position identification data indicating a boundary between a first half part and a second half part of each section, wherein

for an interrupt request generated during reproduction of a predetermined section of said sections, a judgement is made as to whether the interrupt request has been generated in the first half part from a head position to a boundary position of the section during reproduction or generated in the second half part from the boundary position of the section during reproduction to a head position of the next section, and

based on a result of judgement as to generation timing of said interrupt request, as a target section of said interrupt request judged to have been generated in the first half part of said section during reproduction, a previous section of the section during reproduction is specified, or as a target section of said interrupt request judged to have been generated in the second half part of said section during reproduction, said section itself during reproduction is specified.

[Claim 2] A reproducing method of a speech sound recording



medium, for reproducing speech sound data recorded on a predetermined recording medium, wherein

prepared is a speech sound recording medium, comprising: at least, a first region on which a first speech sound data sequence, which is a speech sound data sequence corresponding to one or more sentences made up of a plurality of word sequences to be reproduced and outputted by a predetermined speech sound reproduction means, sectioned into variable-length sections for each of speech sound data respectively divided at pauses in speech sound has been recorded; a second region on which a second speech sound data sequence, which is a speech sound data sequence equivalent to the first speech sound data sequence and made up of slowed speech sound, sectioned into variable-length sections for each of speech sound data equivalent to speech sound data of each section of the first speech sound data sequence has been recorded; a third region on which, in order to reproduce the first and second speech sound data sequences by switching by the speech sound reproducing means, recording position identification data indicating each switchable section by a recording position of each section on the recording medium has been recorded; and a fourth region on which, for each section of the first and second speech sound data sequences, boundary position

identification data indicating a boundary between the first half part and second half part of each section has been recorded,

for a first interrupt request concerning a switching reproduction between said first and second speech sound data sequences generated during reproduction of a predetermined section of said sections, a judgement is made as to whether the first interrupt request has been generated in the first half part from a head position to a boundary position of the section during reproduction or generated in the second half part from the boundary position of the section during reproduction to a head position of the next section,

based on a result of judgement as to generation timing of said first interrupt request, as a target section of the first interrupt request judged to have been generated in the first half part of said section during reproduction, a previous section of the section during reproduction is specified, or as a target section of said first interrupt request judged to have been generated in the second half part of said section during reproduction, said section itself during reproduction is specified, and

speech sound reproduction is restarted from the head of a section of a speech sound data sequence instructed, out of said

first and second speech sound data sequences, by the first interrupt request, which is a section corresponding to the section specified as a target section of said first interrupt request.

[Claim 3] The reproducing method of a speech sound recording medium as set forth in Claim 2, wherein

for a second interrupt request concerning a repeated reproduction generated during reproduction of a predetermined section of said sections, a judgement is made as to whether the second interrupt request has been generated in the first half part from a head position to a boundary position of the section during reproduction or generated in the second half part from the boundary position of the section during reproduction to a head position of the next section,

based on a result of judgement as to generation timing of said second interrupt request, as a target section of the second interrupt request judged to have been generated in the first half part of said section during reproduction, a previous section of said section during reproduction is specified, or as a target section of said second interrupt request judged to have been generated in the second half part of said section during reproduction, said section itself during reproduction is specified, and

in a speech sound data sequence including the section during reproduction out of said first and second speech sound data sequences, speech sound reproduction is restarted from the head of a section specified as a target section of said second interrupt request.

[Claim 4] A reproducing method of a speech sound recording medium, for reproducing speech sound data recorded on a predetermined recording medium, wherein

prepared is a speech sound recording medium, comprising: at least, a first region on which a speech sound data sequence, which is a speech sound data sequence corresponding to one or more sentences made up of a plurality of word sequences to be reproduced and outputted by a predetermined speech sound reproduction means, sectioned into variable-length sections for each of speech sound data respectively divided at pauses in speech sound has been recorded; and a second region on which, for each section of the speech sound data sequence, boundary position identification data indicating a boundary between the first half part and second half part of each section has been recorded,

for an interrupt request concerning a repeated reproduction generated during reproduction of a predetermined section of said sections, a judgement is made as to whether the interrupt

request has been generated in the first half part from a head position to a boundary position of the section during reproduction or generated in the second half part from the boundary position of the section during reproduction to a head position of the next section,

based on a result of judgement as to generation timing of said interrupt request, as a target section of said interrupt request judged to have been generated in the first half part of said section during reproduction, a previous section of the section during reproduction is specified, or as a target section of said interrupt request judged to have been generated in the second half part of said section during reproduction, said section itself during reproduction is specified, and

speech sound reproduction is restarted from the head of a section specified as a target section of said interrupt request.

[Claim 5] A speech sound recording medium on which recorded are: at least, a first region on which a speech sound data sequence, which is a speech sound data sequence corresponding to one or more sentences made up of a plurality of word sequences to be reproduced and outputted by a predetermined speech sound reproduction means, sectioned into variable-length sections for each of speech sound data respectively divided at pauses

in speech sound has been recorded; and

a second region on which, for each section of said speech sound data sequence, boundary position identification data indicating a boundary between the first half part and second half part of each section has been recorded.

[Claim 6] A speech sound recording medium comprising: at least, a first region on which a first speech sound data sequence, which is a speech sound data sequence corresponding to one or more sentences made up of a plurality of word sequences to be reproduced and outputted by a predetermined speech sound reproduction means, sectioned into variable-length sections for each of speech sound data respectively divided at pauses in speech sound has been recorded;

a second region on which a second speech sound data sequence, which is a speech sound data sequence equivalent to said first speech sound data sequence and made up of slowed speech sound, sectioned into variable-length sections for each of speech sound data equivalent to speech sound data of each section of the first speech sound data sequence has been recorded;

a third region on which, in order to reproduce said first and second speech sound data sequences by switching by the speech sound reproducing means, recording position identification data indicating each switchable section by a

recording position of each section on the recording medium has been recorded; and

a fourth region on which, for each section of said first and second speech sound data sequences, boundary position identification data indicating a boundary between the first half part and second half part of each section has been recorded.

[DETAILED DESCRIPTION OF THE INVENTION]

[0001]

[Field of the Invention] The present invention relates to a speech sound recording medium of a medium such as a CD-ROM, a DAT, or the like, on which various types of data including speech sound data is recorded, a reproducing method to reproduce speech sound data recorded in advance on such a speech sound recording medium, and an interrupt handling method for handling an interrupt request to instruct various reproducing methods, which occur during speech sound reproduction.

[0002]

[Prior Arts] Priorly, a variety of teaching materials of speech sound data recorded on a recording medium such as a cassette tape have been provided for the purpose of self-study of a language such as English conversation, the practice of reciting Chinese poems, the study of law, etc. Here, teaching material

for the self-study of English conversation will be described as an example. As a main recording medium according to the prior art, a sequence of English vocal sounds (speech sound data), for example, is recorded on a cassette tape (or a record); and a learner uses this tape teaching material in combination with textbooks. Moreover, for such a teaching material, a variety of levels starting from novice to expert have been provided.

[0003] In addition, in Japanese Patent No. 2581700, proposed is a data recording medium such as a CD-ROM comprising, at least, a first region on which a speech sound data sequence suitable for an expert learner, sectioned into a plurality of sections has been recorded; a second region on which a speech sound data sequence, suitable for a novice learner, made up of equivalent sections corresponding to these respective sections has been recorded; and a third region on which data showing relationships between the corresponding sections of the speech sound data sequences for an expert learner and a novice learner by recorded positions of the respective sections of these sound data sequences in the recording medium has been recorded, and a reproducing method including a switching reproduction, etc., of a data recording medium provided with such a structure.

[0004] On the other hand, in Japanese Unexamined Patent



Publication No. Hei-5-224581, proposed is a technique to realize a repeating function during speech sound reproduction using a language learning CD. Namely, this technique relates to a delay sensing function to sense a delay in an instruction from a learner during normal listening learning, and concretely, a medium on which one-sentence-speech sound data sequences previously index-numbered have been recorded in order is reproduced by use of a predetermined reproduction device. For example, even when such a reproduction device is provided with a repeating function, immediately after a learner has listened to a difficult sentence or a sentence involving pronunciation(s) hard to listen to, etc., a repeating instruction (a machine operation by the learner) is delayed and reproduction of the next sentence starts, therein exists an inconvenience such that the learner cannot listen to the already-listened-to sentence. Therefore, in this prior art, as shown in Fig. 19, index numbers of a segment to be a reproduction target are sequentially switched over at respective points in time of A, B, C, and D, in order, whereby speech sound data of a segment to be a reproduction target is reproduced, however, simultaneously, an index number indicating the previous segment (an index number indicating speech sound data to be repeatedly reproduced) is also retained

by a predetermined retaining means until an elapse of time  $\Delta t$  after next speech sound reproduction is started. In other words, this prior art independently retains an index number of a segment of a reproduction target and an index number of a segment of a repeated reproduction target, and carries out time management so that switching timings of these respective index numbers are differentiated by a predetermined time  $\Delta t$ . By such a construction, a jump signal (data to instruct a reproduction start position) corresponding to an index number maintained at a point in time where a repeat request from the learner was issued is generated to avoid a problem (instruction delay) caused by a delay in a repeat request from the learner.

[0005]

[Themes to be Solved by the Invention] As in the above, in the prior speech sound reproducing method, an instruction delay from an operator (for example, a language learner) such as a repeating function is taken into consideration in advance, and in order to enable repeated reproduction from a sentence intended by an operator, delay sensing of an instruction from the operator is carried out.

[0006] However, such a delay sensing function is a technique on the assumption that predetermined sentences are regularly recorded in order. Accordingly, for example, like

conversations between performers in a movie, conversations in a daily living environment, etc., in reproduction of a series of speech sound data which is made up of a plurality of sentences different in length and in which a randomly occurring non-speech sound period of time such as a condition where no speech sound is reproduced, a condition where only noise is reproduced, a condition where only music (BGM) is reproduced can exist for a long time (for example, one second or longer) between the respective sentences (speech sound data of the respective speakers), it is difficult for the delay sensing function by time management to accurately realize an operation desired by the learner as an operator. Namely, under such circumstances, a case where a repeat target sentence intended by the operator and an actually repeatedly reproduced sentence are different occurs, and consequently, in language learning, etc., there is a problem such that the learner's attention is drawn by a machine operation and a sufficient learning effect is reduced (concentration is lacking).

[0007] Therefore, the present invention has been made to solve the problem as described above, and an object thereof is to provide a speech sound recording medium having a structure to enable reproduction of speech sound data intended by an operator with accuracy even when a series of speech sound data

made up of a plurality of sentences different in length is reproduced, a reproducing method thereof, and an interrupt handling method during speech sound reproduction.

[0008]

[Means for Solving Themes] The present invention relates to a technique to accurately realize reproduction of speech sound data intended by an operator with an instruction delay taken into consideration, even in reproduction of a series of speech sound data, like conversations between performers in a movie, conversations in a daily living environment, etc., which are made up of a plurality of sentences different in length and in which a randomly occurring non-speech sound period of time such as a condition where no speech sound is reproduced, a condition where only noise is reproduced, a condition where only music (BGM) is reproduced can exist between the respective sentences (speech sound data of the respective speakers).

[0009] In particular, on a speech sound recording medium as a reproduction target, at least, one type or more types of speech sound data sequences, which are speech sound data sequences each corresponding to one or more sentences made up of a plurality of word sequences to be reproduced and outputted by a predetermined speech sound reproduction means, sectioned into variable-length sections for each of speech sound data

respectively divided at pauses in speech sound has been recorded. In this specification, speech sound data means a concept and data including sound (speech sound) which is actually audible, and a speech sound data sequence means data itself recorded on a recording medium such as a CD-ROM.

[0010] In detail, a speech sound recording medium according to the present invention comprises: at least, a first region on which a speech sound data sequence, which is a speech sound data sequence corresponding to one or more sentences made up of a plurality of word sequences to be reproduced and outputted by a predetermined speech sound reproduction means, sectioned into variable-length sections for each of speech sound data respectively divided at pauses in speech sound (breaks in vocalization or grammatics to take a breath at the end of a sentence or during a sentence) has been recorded; and a second region on which, for each section of the speech sound data sequence, boundary position identification data indicating a boundary between a first half part and a second half part of each section.

[0011] Herein, the speech sound recording medium may comprise a plurality of regions on which a plurality of types of speech sound data sequences have been recorded. Namely, for a delay sensing function during a speech sound reproduction, in such

a case, as indicated in the above-described Japanese Patent No. 2581700, where a switchover reproduction of a plurality of types of speech sound data is possible, a speech sound recording medium according to the present invention (second embodiment according to a speech sound recording medium) comprises: at least, a first region on which a first speech sound data sequence, which is a speech sound data sequence corresponding to one or more sentences made up of a plurality of word sequences to be reproduced and outputted by a predetermined speech sound reproduction means, sectioned into variable-length sections for each of speech sound data respectively divided at pauses in speech sound has been recorded; a second region on which a second speech sound data sequence, which is a speech sound data sequence equivalent to the first speech sound data sequence and made up of slowed speech sound, sectioned into variable-length sections for each of speech sound data equivalent to speech sound data of each section of the first speech sound data sequence has been recorded; a third region on which, in order to reproduce the first and second speech sound data sequences by switching by the speech sound reproducing means, recording position identification data indicating each switchable section by a recording position of each section on the recording medium has

been recorded; and a fourth region on which, for each section of the first and second speech sound data sequences, boundary position identification data indicating a boundary between the first half part and second half part of each section has been recorded.

[0012] In addition, a speech sound recording medium of any of the above-described constructions is characterized in that boundary position identification data indicating a boundary between a first half part and a second half part split according to the state of speech sound (for each cohesiveness of speech sound) within each section has been recorded at a predetermined region of the same speech sound recording medium.

[0013] On the other hand, in a speech sound reproducing method according to the present invention, in response to an interrupt request to instruct various reproducing methods generated during speech sound reproduction, a target of the interrupt request is specified by making use of the above-described boundary position identification data set in advance for each section. Herein, a reproduction instruction by an interrupt request includes, at least, a repeating function, a switchover reproduction function for reproducing by switching a plurality of speech sound data sequences stored in respective regions of the speech sound recording medium, or a combination of these.

[0014] In a case where the above-described repeating function is realized, in a reproducing method of a speech sound recording medium according to the present invention (first embodiment according to a reproducing method), first, prepared is a speech sound recording medium comprising: at least a first region on which a speech sound data sequence, which is a speech sound data sequence corresponding to one or more sentences made up of a plurality of word sequences to be reproduced and outputted by a predetermined speech sound reproduction means, sectioned into variable-length sections for each of speech sound data respectively divided at pauses in speech sound has been recorded; and a second region on which, for each section of the speech sound data sequence, boundary position identification data indicating a boundary between the first half part and second half part of each section split according to the state of speech sound has been recorded. And, when an interrupt request concerning a repeated reproduction from an operator is issued during reproduction of a predetermined section among a plurality of recorded sections, generation timing of this interrupt request is judged. In detail, a judgement is made as to whether the interrupt request has been generated in the first half part from a head position to a boundary position of the section during reproduction or



generated in the second half part from the boundary position of the section during reproduction to a head position of the next section. Here, if the above-described interrupt request is judged to have been generated in the first half part of the section during reproduction, a previous section of the section during reproduction is specified as a target section of the interrupt request. On the other hand, if the above-described interrupt request is judged to have been generated in the second half part of the section during reproduction, the section itself during reproduction is specified as a target section of the interrupt request.

[0015] In the first embodiment according to a reproducing method, a repeating function is realized by restarting a speech sound reproduction from the head of a section (section specified as a target section intended by an operator) specified by interrupt handling as in the above.

[0016] On the other hand, in a case where a plurality of speech sound data sequences recorded on the recording medium is reproduced, in a reproducing method of a speech sound recording medium according to the present invention (second embodiment according to a reproducing method), first, prepared is a speech sound recording medium comprising: at least, a first region on which a first speech sound data sequence, which is a speech

sound data sequence corresponding to one or more sentences made up of a plurality of word sequences to be reproduced and outputted by a predetermined speech sound reproduction means, sectioned into variable-length sections for each of speech sound data respectively divided at pauses in speech sound has been recorded; a second region on which a second speech sound data sequence, which is a speech sound data sequence equivalent to the first speech sound data sequence and made up of slowed speech sound, sectioned into variable-length sections for each of speech sound data equivalent to speech sound data of each section of the first speech sound data sequence has been recorded; a third region on which, in order to reproduce the first and second speech sound data sequences by switching by the speech sound reproducing means, recording position identification data indicating each switchable section by a recording position of each section on the recording medium has been recorded; and a fourth region on which, for each section of the first and second speech sound data sequences, boundary position identification data indicating a boundary between the first half part and second half part of each section split according to the state of speech sound has been recorded. And, when an interrupt request concerning a switchover reproduction between the first and second speech sound data sequences is

generated during reproduction of a predetermined section of these sections, generation timing of the interrupt request is once judged. Namely, a judgement is made as to whether this interrupt request has been generated in the first half part from a head position to a boundary position of the section during reproduction or generated in the second half part from the boundary position of the section during reproduction to a head position of the next section, and based on a result of this judgement, a section to be an interruption target is specified. If this interruption request is judged to have been generated in the first half part of the section during reproduction, a previous section of the section is specified, and if the interruption request is judged to have been generated in the second half part of the section during reproduction, the section itself is specified.

[0017] In the second embodiment according to a reproducing method, when a section as an instruction target intended by an operator is specified by interrupt handing as in the above, speech sound reproduction is restarted from the head of a section, which is a section corresponding to a section specified as a target section of the interrupt request, of a speech sound data sequence instructed, out of the first and second speech sound data sequences, by the above-described

interruption request, whereby a switchover reproducing function is realized.

[0018] Herein, even in a reproducing method of a speech sound recording medium where switchover reproduction is possible as in the above, it is possible to further provide various functions including the above-described repeating reproduction.

[0019] In addition, a speech sound recording medium according to the present invention includes a data recording medium which can be reproduced by a portable CD player and a personal computer with a CD-ROM driver or the like. Accordingly, in order to be reproduced by such various devices, image data and text data, etc., which can be reproduced in a manner corresponding to the reproduction timing of respective speech sound data may be included.

[0020]

[Preferred Embodiment] Hereinafter, an embodiment of the present invention will be described in detail with reference to Fig. 1 through Fig. 18. Herein, identical symbols are used for identical parts and description thereof will be omitted.

[0021] (First embodiment according to a speech sound recording medium) First, a basic structure of speech sound data to be recorded on a speech sound recording medium according to the

present invention will be schematically described by use of Fig. 1.

[0022] Data to be recorded on a speech sound recording medium according to the present invention is, like conversations between performers in a movie, conversations in a daily living environment, etc., a series of speech sound data which is made up of a plurality of sentences different in length and in which a randomly occurring non-speech sound period of time such as a condition where no speech sound is reproduced, a condition where only noise is reproduced, a condition where only music (BGM) is reproduced can exist between the respective sentences (speech sound data of the respective speakers). Accordingly, a first embodiment (basic construction) according to this speech sound recording medium comprises, at least, a first region on which a speech sound data sequence, which is a speech sound data sequence corresponding to one or more sentences made up of a plurality of word sequences to be reproduced and outputted by a predetermined speech sound reproduction means, sectioned into variable-length sections (hereinafter, referred to as segments) for each of speech sound data respectively divided at pauses in speech sound has been recorded.

[0023] Generally, in an English conversation of a native

speaker, since one sentence is vocalized in approximately three seconds, it is appropriate to respectively construct variable-length segments to make up a speech sound data sequence, as shown in Fig. 1(a), (b), or (d), by setting pauses in speech sound to determine segments to make up a speech sound data sequence to be recorded between respective sentences. Herein, although sentences in a conversation can include an extremely short sentence, as shown in Fig. 1(c), this sentence also forms one segment. On the other hand, as shown in Fig. 1(e), if the sentence is extremely long, a pause of speech sound comes before a conjunction, relative clause, etc., therefore, in a case of the sentence as shown in Fig. 1(e), it is appropriate to make up the sentence of two consecutive segments. Accordingly, segments of a speech sound data sequence to be recorded mean recording units of speech sound data divided based on breaks according to speech sound (breathing positions) or some linguistic (grammatics) breaks.

[0024] In addition, the speech sound recording medium further comprises a second region on which, in order to appropriately handle an interrupt request which is possibly generated during reproduction of special speech sound data as described above, boundary position identification data indicating a boundary between a first half part and a second half part of the

respective segments divided at pauses in speech sound as described above has been recorded. Concretely, this boundary position identification data is recorded, for example, on a directory region of the speech sound recording medium or a header portion of each segment.

[0025] A boundary to be set for each of the divided segments is set, as shown in Fig. 1(a), (b), or (d), to a first break S1, S2, or S4 with some cohesiveness of speech sound (namely, S1, S2, S3, or S4 is set for each section according to the state of speech sound, whereby each section is split into a first half part and a second half part.) Herein, as shown in Fig. 1(c), if the sentence is extremely short, the boundary position S3 may be set so as to coincide with the end position of the divided segment. In addition, as shown in Fig. 1(e), if one sentence is divided into a plurality of segments, boundary positions S5 and S6 each indicating a boundary to split each segment into a first half part and a second half part judged to be cohesiveness of speech sound are set, respectively.

[0026] In particular, speech sound reproduction of a speech sound recording medium according to this first embodiment is carried out for each of recorded segments in order. However, if an interrupt request is made from an operator such as a repeat instruction during such speech sound reproduction, generation

timing of this interrupt request is judged, and based on a result of this judgement, a segment to be a target of the interrupt request is specified. Concretely, it is judged as to whether this interrupt request has been generated in the first half part from a head position of a segment during reproduction to a boundary position set as described above or generated in the second half part from the boundary position of the segment during reproduction to a head position of the next segment to be reproduced next. Herein, if it is judged that the above-described interrupt request has been generated in the first half part of a segment during reproduction, a previous segment (a segment whose speech sound reproduction has already finished) of the segment during reproduction is specified as a target segment of the interrupt request. On the other hand, if it is judged that the above-described interrupt request has been generated in the second half part of the segment during reproduction, the segment itself during reproduction is specified as a target segment of the interrupt request.

[0027] This reproducing method realizes a repeating function by restarting speech sound reproduction from the head of a segment (a segment specified as an instruction target by the operator) specified as such.



[0028] (Second embodiment according to a speech sound recording medium) Next, a second embodiment of a speech sound recording medium according to the present invention is characterized in that at least two types of speech sound data sequences have been recorded. Namely, a first speech sound data sequence is composed of speech sound data of English spoken by a native speaker at a natural speed, and this speech sound data sequence is divided into a plurality of variable-length segments at pauses in speech sound (breaks in vocalization or grammatics to take a breath at the end of a sentence or during a sentence), as described above. Although a second speech sound data sequence has meaning contents equivalent to the contents of the above-described first speech sound data sequence, this is of different speech sound data and is, for example, English speech sound data spoken at a slower speed by separating the data word by word. Moreover, this second speech sound data sequence is also made up of a plurality of variable-length segments.

[0029] The important thing here is although the above-described first and second speech sound data sequences are sectioned respectively into pluralities of variable-length segments, these sequences correspond to each other in their meaning contents segment by segment. For example, when a t-th (the

621<sup>st</sup> in Fig. 1(a)) segment of the first speech sound data sequence is "It's not much of a problem." spoken by a native speaker, a t-th segment of the second speech sound data sequence becomes "It is not much of a problem." spoken by separating the data word by word. However, having contents corresponding to the second speech sound data sequence and being made up of different speech sound data indicates that the data has an identical meaning in language and has a different vocalization. [0030] Furthermore, a speech sound recording medium according to the present second embodiment is characterized in that recording position identification data has been recorded in a predetermined region. This recording position identification data is, at least, data indicating in which position of the speech sound recording medium each segment of the first and second speech sound data sequences has been recorded. Accordingly, in which position of the medium the segment of the second speech sound data sequence "It is not much of a problem." corresponding to the t-th segment of the first speech sound data sequence "It's not much of a problem." can be recognized by this recording position identification data.

[0031] As a result, the first and second speech sound data sequences are recorded with no relation to each other but having

a fixed relation, and the respective speech sound data sequences are organically combined using the segments as units. Namely, the first and second speech sound data sequences are paired with each other, and what correlates these segment by segment is the recording position identification data. Moreover, in the present embodiment, the recording position identification data is recorded in a directory region of the speech sound recording medium and contains, at least, data concerning a head position of each segment. Moreover, in this embodiment, boundary position identification data is recorded on a header portion of each segment.

[0032] In a reproducing method of a speech sound recording medium (second embodiment) having a structure as in the above, as well, speech sound reproduction is carried out for each of recorded segments in order, and in particular, this reproducing method is characterized in that a reproduction switchover from a first speech sound data sequence to a second speech sound data sequence (or a reproduction switchover from a second speech sound data sequence to a first speech sound data sequence) is possible. Moreover, this reproduction switchover operation is carried out for the segments as units. For example, when a reproduction instruction to the second speech sound data sequence is inputted during reproduction of

a t-th segment of the first speech sound data sequence (generation of an interrupt request), a corresponding t-th segment of the second speech sound data sequence is read out based on recording position identification data, and speech sound reproduction of the corresponding segment is executed. In addition, in contrast thereto, a reproduction switchover from the second speech sound data sequence to the first speech sound data sequence is also, similar to the above-described reproduction switchover operation, carried out for each segment unit.

[0033] Moreover, in the above-described switchover operation, specification of a segment to be a switching target is carried out by judging interruption timing as explained in the reproducing method of a speech sound recording medium according to the above-described first embodiment. In addition, in this reproducing method, in addition to the above-described reproduction switchover operation, various modifications including a repeated reproduction are possible. As a representative thereof, a so-called return instruction exists. Namely, when a return instruction is inputted after reproduction is temporally interrupted based on a stop order during reproduction, a reading-out position of speech sound data is returned by an instructed amount, whereby reproduction

of speech sound data to further satisfy an operator's wish is carried out.

[0034] (Third embodiment of a speech sound recording medium) Furthermore, a third embodiment of a speech sound recording medium according to the present invention will be described. A speech sound recording medium according to the present third embodiment differs from the speech sound recording medium according to the above-described second embodiment in that, in addition to first and second speech sound data sequences, a third speech sound data sequence is further recorded on the same speech sound recording medium.

[0035] The important thing here is that the above-described three speech sound data sequences are sectioned into segment groups with one or more variable-length segments of the first and second speech sound data sequences integrated. In other words, one segment group contains one or more segments of the first and second speech sound data sequences, and accordingly, one segment group of the third speech sound data sequence is paired with one or more segments of the first and second speech sound data sequences. In particular, this construction is on the assumption that one sentence is sectioned into a plurality of segments as shown in Fig. 1(e).

[0036] In addition, in a speech sound recording medium of the

present third embodiment, recording position identification data recorded in a predetermined region also includes data indicating recording positions of the contents of the above-described third speech sound data sequence as well for each segment group. Accordingly, the first, second, and third speech sound data sequences and recording position identification data are recorded on the medium while having a fixed relationship with each other, and the respective speech sound data sequences are organically combined using the segments or segment groups as units. In this third embodiment as well, recording position identification data is recorded in a directory region of the same speech sound recording medium, and contains data concerning head positions of segments of the respective speech sound data sequences. Moreover, in this embodiment as well, boundary position identification data is recorded on a header portion of each segment.

[0037] A reproducing method of a speech sound recording medium (third embodiment) having a structure as in the above is basically the same as in the above-described second embodiment, however, this differs from that of the second embodiment in that, in addition to a reproduction switchover between the first and second speech sound data sequences, a reproduction switchover operation is carried out between these first and

second speech sound data sequences and third speech sound data sequence, as well.

[0038] For example, if "It's not much of a problem." by a native speaker cannot be heard during reproduction of the first speech sound data sequence, by switching over reproduction from the first speech sound data sequence during reproduction to the second speech sound data sequence, "It is not much of a problem." spoken slowly by separating the data word by word can be heard. And, when the meaning and the grammar of this are desired to be known in Japanese, the reproduction may be further switched from the speech sound data sequence during reproduction to the third speech sound data sequence. As a matter of course, in this reproducing method as well, it is needless to say that application is possible so that a return instruction and a stop order as explained in the reproducing method of a speech sound recording medium according to the above-described second embodiment can be used in combination. In addition, in this reproducing method as well, switching reproduction and repeated reproduction are possible, and segments as targets of these interrupt requests are specified similarly as with the case of the above-described first embodiment.

[0039] (Fourth embodiment of a speech sound recording medium)  
A fourth embodiment of a speech sound recording medium

according to the present invention is basically similar to that of the above-described third embodiment, however, this is different mainly in that, in addition to first and second speech sound data sequences, a text data sequence is recorded. This text data sequence corresponds to text data having contents corresponding to the first or second speech sound data sequence, and corresponds to text data corresponding to English (speech sound) spoken by a native speaker, for example.

[0040] This text data sequence is also sectioned into segments corresponding to the respective segments of the first and second speech sound data sequences. In addition, in a speech sound recording medium according to this fourth embodiment as well, recording position identification data includes data of recording positions of this text data sequence, for respective segments of the respective speech sound data sequences, concerning head positions of these, and the data is recorded in a directory region of the same speech sound recording medium. In addition, on a header portion of each segment, boundary position identification data is recorded. Accordingly, the first and second speech sound data sequences and the text data sequence correspond to each other segment by segment. Moreover, in a speech sound recording medium of the present fourth embodiment, if the third speech sound data sequence of



the above-described third embodiment is added as recording data, one or more segments of the first and second speech sound data sequences and text data sequence also correspond to one segment group of the third speech sound data sequence. Furthermore, in the present fourth embodiment as well, the above-described recording position identification data includes a head position of each segment, and is recorded in a directory region of the same speech sound recording medium. In addition, boundary position identification data is recorded on a header portion of each segment.

[0041] A reproducing method of a speech sound recording medium (fourth embodiment) having a structure as in the above is also basically similar to that of the above-described second embodiment, however, this is different in that, the text data sequence is displayed during reproduction of the first or second speech sound data sequence.

[0042] For example, when a segment of the first speech sound data sequence "It's not much of a problem." is being reproduced, "It's not much of a problem." or "It is not much of a problem." is displayed in a predetermined display part. It is not necessary for this display to chronologically completely synchronize with the speech sound data sequence during reproduction, and the text may be displayed with a slight delay

or may be displayed slightly ahead of time. In addition, in this reproducing method as well, a switching reproduction and a repeated reproduction are possible, and segments as targets of these interrupt requests are specified similarly as with the reproducing method of a speech sound recording medium according to the first embodiment.

[0043] Next, a detailed structure of a speech sound recording medium according to the present invention will be hereinafter described in detail by use of Fig. 2 through Fig. 5.

[0044] Fig. 2 is a view for explaining respective speech sound data sequences A, B, and C when the above-described third embodiment is applied for self-study of English conversation, and recording contents thereof. In this view, a speech sound data sequence A is a data sequence (first speech sound data sequence) spoken by a native speaker, which is composed of a plurality of segments 621 and 622. A speech sound data sequence B is a data sequence (second speech sound data sequence) of English spoken slowly by separating the data word by word, which is also composed of a plurality of segments corresponding to the segments 621 and 622 of the previous speech sound data sequence A made up of English words or phrases. In addition, a speech sound data sequence C is a data sequence (third speech sound data sequence) for Japanese explanation,

and segment groups contained in this speech sound data sequence C respectively correspond to respective segments 621 and 622 of the respective speech sound data sequences A and B.

[0045] In addition, Fig. 3 is a table for explaining a relationship between time and capacity of each segment in the mode as shown in Fig. 2. In this table, one second corresponds to a capacity of 6 kilobytes. For example, in the segment 621 of the speech sound data sequence A, a vocalization time for "it's" is 0.2 seconds, its capacity is 1.2KB (kilobytes), a vocalization time for "not" is 0.1 seconds, its capacity is 0.6KB (kilobytes), a vocalization time for "much of a" is 0.4 seconds, its capacity is 2.4KB (kilobytes), and a vocalization time for "problem" is 0.3 seconds, and its capacity is 1.8KB (kilobytes). A vocalization time of the whole segment 621 is 2.0 seconds, and its capacity is 1.8KB (kilobytes). Figures in parentheses in the table are values expressing boundary positions by reproducing times. Namely, in a case of this speech sound data sequence A, as shown in Fig. 1(a), since a boundary position is set after "not," this indicates that a boundary to separate the speech sound data sequence A into a first half part and a second half part is set at a position 0.3 seconds after a reproduction start of the segment 621. In addition, although a boundary is similarly shown for other

speech sound data sequences B and C as well, in this embodiment of Fig. 3, no boundary position identification data is set for the speech sound data sequence C.

[0046] Furthermore, Fig. 4 is a table for explaining recording contents of a directory region in the modes as shown in Fig. 2 and Fig. 3. In this table, one segment of the directory region is formed of  $9 \times 3 = 27$  bytes (B). Speech sound data sequences A, B, and C correspond to the speech sound data sequences A, B, and C of Fig. 2, respectively. In addition, C of one byte indicates an attribute, wherein  $C = 0$  means a speech sound data sequence A, and  $C = 64$  means a speech sound data sequence B. In addition,  $C = 128, 129$  means a speech sound data sequence C, and in particular, when  $C = 129$ , that is to say "10000001" according to the bit expression (8 bits), this indicates that it is the same explanation target as that of the previous segment (this indicates that it belongs to the same segment group to be an explanation target of the speech sound data sequence C, and for example, a case of the segments 801 and 802 of Fig. 1(e) corresponds hereto.)

[0047] Position data M, S and B (one byte each) is a parameter showing the positions in a CD-ROM that has been standardized in the industry. That is to say M indicates minutes, S indicates seconds, and B indicates blocks, respectively.

Moreover, one block is made up of 2048 bytes. Thus, 75 blocks form an amount for one second. Accordingly, the maximum numbers are  $M = 59$ ,  $S = 59$ , and  $B = 74$ . SB of the next two bytes indicates the start byte and LLL of the following three bytes indicates the whole length of each segment. Here, minutes and seconds are used as parameters indicating position, because the CD-ROM was originally developed for music recording, and therefore, the recorded position is represented as a length of a time from the start. Thus, in a case where the speech sound recording medium employs a CD-ROM, minutes and seconds are not at all related to the period of time of reproduction, but rather this data simply represents the recorded position.

[0048] As a result, as for "It's not much of a problem." of segment 621 of the speech sound data sequence A, for example, speech sound data of English spoken by a native speaker is recorded starting from the 826<sup>th</sup> byte of the 0 minutes 11 seconds 3 blocks with a length of 6000 bytes, as for a corresponding segment of the speech sound data sequence B, English spoken slowly by separating the data word by word is recorded starting from the 2026<sup>th</sup> byte of the 0 minutes 11 seconds 3 blocks with a length of 7400 bytes, and as for a segment groups of the speech sound data sequence C, Japanese explanation is recorded starting from the 1282<sup>nd</sup> byte of 0 minutes 11 seconds 6 blocks

with a length of 72000 bytes. Here, segment numbers 621, 622, etc., are not in the memory but correspond to the addresses thereof. In addition, recording position identification data showing a relationship between the segments is included in this directory region.

[0049] In greater detail, based on the recording contents of a directory region shown in Fig. 4, in a region from the 826<sup>th</sup> byte to the  $826 + 6,000 - 1 = 6,825^{\text{th}}$  byte of the 0 minutes 11 seconds 3 blocks of the same speech sound recording medium, a speech sound data sequence whose segment is 621 and whose attribute C is 0, that is, data equivalent to "It's not much of a problem." spoken by a native speaker is recorded. In addition, in a region from the 2,026<sup>th</sup> byte to the  $2,026 + 17,400 - 1 = 19,425^{\text{th}}$  byte of the 0 minutes 11 seconds 3 blocks of the same speech sound recording medium, a speech sound data sequence whose segment is 621 and whose attribute C is 64, that is, data equivalent to "It is not much of a problem." spoken slowly by separating the data word by word is recorded. Furthermore, in a region from the 1,282<sup>nd</sup> byte to the  $1,282 + 72,000 - 1 = 73,281^{\text{st}}$  byte of the 0 minutes 11 seconds 6 blocks of the same speech sound recording medium, a speech sound data sequence whose segment is 621 and whose attribute C is 128, that is, data equivalent to Japanese explanation is recorded.

[0050] As such, if the directory region as shown in Fig. 4 is provided, the respective speech sound data sequences shown in Fig. 2 can be recorded with reproducing times and capacities as shown in Fig. 3.

[0051] Next, boundary position identification data showing a border between a first half part and a second half part of the respective segments 621 and 622 is recorded on a header portion of a variable-length segment shown in Fig. 5(a), for example. This header portion is, as shown in Fig. 5(b), a region of 32 bytes made up of, from the head: a 1-byte region (1B) to show the presence or absence of text data and image data, etc.; a 5-byte region, which is a region prepared for a speech sound data sequence A, composed of 1-byte data showing a data sequence type (data to distinguish between speech sound data sequences A, B, etc.), 3-byte data showing its data length, and 1-byte data showing a boundary position; a 5-byte region, which is a region prepared for a speech sound data sequence B, composed of 1-byte data showing a data sequence type, 3-byte data showing its data length, and 1-byte data showing a boundary position; a 4-byte region (4B), which is a region prepared for a speech sound data sequence C, composed of 1-byte data showing a data sequence type and 3-byte data showing its data length; a 4-byte region (4B), which is a region prepared for a text data

sequence D, composed of 1-byte data showing a data sequence type and 3-byte data showing its data length; a 6-byte region (6B), which is also a region prepared for a text data sequence D, composed of 3-byte data showing an address and 3-byte data showing its data length; a 4-byte region (4B) prepared for another data sequence (type E); and a preliminary 3-byte region (3B).

[0052] In an embodiment shown in this specification, another data sequence (type E) is not used. In addition, in this embodiment, boundary position identification data is set only for the speech sound data sequences A and B. In the 1-byte region showing a boundary position, a quotient obtained by dividing, in terms of data after a compression, a byte number from the head of segments by 256 is set. For example, if a speech sound sampling rate of track A (a virtual track equivalent to a speech sound data sequence A) is provided as 128 kilobit/second (16 kilobytes/second), a boundary position within a range from the head to 4 seconds can be expressed by this 1-byte region, and its resolution results in 0.016 seconds. In addition, in a case where the speech sound sampling rate is 32 kilobit/second (4 kilobytes/second), a boundary position within a range from the head to 16 seconds can be expressed by this 1-byte region, and its resolution results in 0.064



seconds, thus performance sufficient for practical use can be obtained.

[0053] Next, a basic reproducing method of a speech sound recording medium according to the present invention will be described by use of Fig. 6 through Fig. 10.

[0054] First, Fig. 6 is a perspective view showing a whole construction of a reproduction unit to realize a reproducing method of a speech sound recording medium according to the present invention. As can be understood from this drawing, as well, the speech sound recording medium is a CD-ROM which can be reproduced by a portable CD player (reproduction unit body 2), for example, and this reproduction unit body 2 is remotely controlled by a cord-connected handset 8. In this handset 8, a display part 21 such as a liquid crystal display (LCD) to display, at least, a segment number during reproduction and various control button groups 240 are provided. In addition, an operator can listen to speech sound data reproduced by the reproduction unit body 2 via an earphone 13.

[0055] In addition, Fig. 7 is a block diagram showing a construction of the reproduction unit shown in Fig. 6. As shown in this drawing, a CD-ROM, which is the speech sound recording medium 1, is set in a reproduction mechanism 20. The reproduction mechanism 20 is connected to a CPU 5 via a disk

interface (I/F) 3 and a bus 4. In addition, a ROM 6 of 32 kilobytes (KB), for example, for storing a program and a RAM 7 of 256 kilobytes, for example, for temporarily storing a directory or speech sound data sequence are connected to the bus 4. Furthermore, the bus 4 is connected to a handset interface (I/F) 9 for sending and receiving data to and from the handset 8 for manual operation and is connected to a D/A converter 12 that is connected to an external terminal 11 and handset 8 via an amplifier for speech sound output (AMP) 10. Here, an earphone 13 is connected to the handset 8 as described above.

[0056] Figs. 8 (a) and (b) are diagrams showing memory allocations of the ROM 6 and RAM 7, respectively. As shown in Fig. 8(a), a program of 32 kilobytes is stored in the ROM 6. On the other hand, as shown in Fig. 8(b), to the RAM 7, a buffer of  $(50 + 50 =) 100$  kilobytes (equivalent to 50 blocks), a directory of  $(75 + 75 =) 150$  kilobytes, and a system area of 6 kilobytes are allocated. Accordingly, a speech sound data sequence of 50 blocks is always stored in the RAM 7, and a directory (only the portion of speech sound data sequence A takes up approximately 30 minutes) for  $150 \text{ kilobytes} + 27 \equiv 5555$  segments is also stored in the RAM 7.

[0057] Here, a CD-ROM is used as the speech sound recording

medium in the above-described detailed example, and the capacity of a representative CD-ROM is 552 megabytes (MB). Units of minutes, seconds, and blocks are used to represent the address in the CD-ROM. Moreover, one block is formed of 2048 bytes, 75 blocks form one second, and 60 seconds form one minute, therefore, the CD-ROM holds the maximum address value of 59 minutes 59 seconds 74 blocks. On the other hand, the capacity of this CD-ROM becomes  $(2048 \times 75 \times 60 \times 60 =)$  552.96 megabytes. The first two seconds thereof are used for formatting the CD-ROM and cannot be used by the user and thereby, the maximum capacity becomes 552.6528 MB to be precise. In the case where a directory is allocated to a portion corresponding to the first 20 seconds out of the maximum capacity, a directory capacity of 3 megabytes can be secured in the CD-ROM.

[0058] Next, a calculation example concerning the capacity is shown.

[0059] A speech sound sampling rate is 16 kilo-samples/second according to an ADPCM system and 3 bits are used for one sample. This presumption leads to 48 kilobits/second (Kbit/s) and accordingly, 6 kilobytes/second (KB/s), and therefore, it is necessary for the transfer rate to be adjusted to this rate. Here, in the case of 16 kilo-sample/second, f characteristics

up to 8 kilohertz (KHz) are exhibited. Accordingly, consonants can be recorded sufficiently. According to the above-described presumption, a capacity of (6 kilobytes  $\times$  3600 seconds  $=$ ) 21.6 megabytes is required for recording sound for one hour. In general, 552 megabytes are recorded on one CD-ROM, including error corrections. The CD-ROM, excluding the directory region, can use 549 megabytes as a data region in which speech sound data is stored. Accordingly, it becomes possible to record speech sound data for (549  $\div$  21.6  $=$ ) 25 hours 24 minutes. Thus, in the case where the CD-ROM is used for the study of English conversation, supposing that a one hour story spoken by a native speaker at a natural speed is provided, the portion of the story that is spoken slowly by separating the data word by word would take four hours, four times longer than the natural speed. Thus, the total hours become 20 hours even in the case where 15 total hours are required for the explanation portion.

[0060] Next, the method of dividing the one hour conversation portion into several segments (sections) is considered and then presuming that each second is made up of four sections on average, the one hour would be divided into 14400 segments. 36 bytes are required for a directory for one segment, and therefore, approximately 518 kilobytes are required in total

which can be sufficiently stored in the above-described storage place for directories of 3 megabytes, and therefore, all of the directories for the one hour story can be stored.

[0061] Next, Fig. 9 is a plan view showing a construction in detail of the handset 8 and earphone 13 shown in Fig. 6 and Fig. 7. The front side of handset 8 is provided with a segment display part 21 (LCD) for displaying the segment number, a display part 22 for displaying contents of instruction at that point in time by utilizing the LED, a mode switch 81 for switching over the reproduction mode (showing a reproducing condition in a case where repeated reproduction is carried out), instruction switches 23A, 23B and 23C for instructing the respective reproductions of speech sound data sequences A, B, and C, an ADV button 24A for instructing the function, a REV button 24B for instructing the return of reproduction, a REP button 25A for instructing the repeated reproduction, and a STOP button 25B for instructing the stoppage of the reproduction, and a volume switch 28 for adjusting the volume outputted from the earphone 13. In addition, the earphone 13 is connected to the handset 8 via a cord 26, and the handset 8 is connected to the reproduction unit body 2 via a cord 27.

[0062] Next, an example of a basic reproduction sequence of a reproducing method of a speech sound recording medium

according to the present invention will be explained. Figs. 10 are views for explaining this basic reproduction sequence. When the portion of speech sound data sequence A, that is, the English of the native speaker is continuously heard, speech sound data of recorded segments is reproduced in order as shown in this Fig. 10(a). In this case, a button 23A shown in Fig. 9 is pressed so that the reproduced sound is heard as is.

[0063] On the other hand, in a reproduction sequence shown in Fig. 10(b), first, speech sound data sequence A is reproduced from segment 621, in order. Here, when segment 623 cannot be heard well, an operator immediately presses the STOP button 25 B to once interrupt speech sound reproduction. At this time, the segment number of the segment display part 21 has become 624. Thus, the REV button 24B is pressed only once. Here, when the REV button 24B is held down, the segment continuously moves backwards and every time the REV button 24B is pressed, the segment number displayed in the display part 21 moves backwards by one. When the REV button 24B is pressed once under the condition where the segment is 624, the segment number displayed in the play part 21 becomes 623.

[0064] Subsequently, when the switch 23B is pressed (a switching reproduction instruction), segment 623 of speech sound data sequence B corresponding to the segment 623 of speech

sound data sequence A is reproduced, and speech sound is heard spoken slowly by separating the data word by word. When the unit is left as is, it proceeds to a reproducing operation of segment 624 through speech sound data sequence B. Furthermore, when this segment 624 of speech sound data sequence B is heard, the STOP button 25B is pressed and the REV button 24 is again pressed so that the segment number of the display part 21 becomes 623. In this condition, when the switch 23C is pressed, the Japanese explanation of speech sound data sequence C is reproduced. Some portions of the explanation relate to phrases made up of several words, and therefore, several segment numbers may be covered by one explanation. When the lowest level bit of the byte showing attribute C is 1, this segment (where attribute C is 129) shows the attribute with a meaning of the same explanation target as the previous segment number of the same speech sound data sequence. When such a segment is encountered, it can be skipped.

[0065] Next, another embodiment of a speech sound recording medium according to the present invention is described by use of Fig. 11 through Fig. 13.

[0066] Fig. 11 is a view showing another construction example of a directory region of a speech sound recording medium according to the present invention. In the directory shown

in this view, one segment is formed of 15 bytes. That is to say, attribute C of one byte, segment number SS of two bytes, suffix number N of one byte, minute M of one byte, second S of one byte, block B of one byte, and segment length LLL of three bytes add up to 12 bytes.

[0067] In addition, in terms of the 8 bits of attribute C, when the first bit (top level bit) is 1, it indicates the start of the segment and when the first bit is 0, it indicates another condition. In addition, when a value expressed by the second and third bits of this attribute C is 0, it indicates that the segment is of speech sound data sequence A; when the value is 1, it indicates that the segment is of speech sound data sequence B; and when the value is 2, it indicates that the segment is of speech sound data sequence C. When the fourth bit is 0, it indicates there is no suffix; and when the fourth bit is 1, it indicates there is a suffix. When the fifth bit is 1, it indicates the segment is associated with a segment group the same as that of the previous segment; and when the fifth bit is 0, the segment is not associated with the same segment group.

[0068] Fig. 12 is a view for explaining respective speech sound data sequences A, B, and C being divided into segments corresponding to the directory region shown in Fig. 11. Thus,



this differs from the example shown in Fig. 2 in the point that the data sequences are divided into segments with suffixes.

[0069] In addition, Fig. 13 is, as shown in the above, a view showing a relationship between the values of the second and third bits of attribute C which has 8 bits and the segments corresponding to divided respective sections.

[0070] Furthermore, another embodiment of a speech sound recording medium according to the present invention is described by use of Fig. 14 and Fig. 15. Thus, this differs from embodiments of Fig. 11 through Fig. 13 described in the foregoing in the point wherein text data is recorded and reproduced in addition to speech sound data.

[0071] Fig. 14 is a view showing a directory structure of the same. In this view, the directory differs from the directory of Fig. 11 described in the foregoing in the point wherein data concerning text data sequence D is added to the second and third bits of attribute C. When this bit is 3, it means that text data sequence D is recorded on the speech sound recording medium in the form of a predetermined code.

[0072] Fig. 16 is a plan view showing a handset construction used in a reproduction unit for a speech sound recording medium according to this embodiment. Here, this handset of Fig. 15 differs from the handset of Fig. 9 in the point wherein text

data sequence D can also be displayed. In detail, the difference is in a point wherein the handset 8 is provided with a text display part 210 formed of an LCD in addition to an operation button 23D for instructing a text display. In addition, in this embodiment, a text display button for turning on and off the display of text is further provided. In such a handset 8, setting as to whether the text data sequence D is displayed or is not displayed in the text display part 210 is possible.

[0073] Next, the characteristics of the effects of another embodiment shown in Fig. 11 through Fig. 13 are described.

[0074] In this embodiment, the speed of the appearance of text data sequence D is controlled in accordance with the speed of speech sound data sequence A being spoken, that is to say, the length of LLL, during the reproduction of speech sound data sequence A. Thus, text begins to appear on display part 210 and completely finish outputting during the time from the start of and to the end of the speech sound of the segment (the text is displayed in complete synchronization with the speech sound).

[0075] Next, text data sequence D is displayed in synchronization with the length of the speech sound data sequence B during speech sound data sequence B's reproduction,

and at this time the speech sound is reproduced for a period of time several times longer than the time of reproduction of speech sound data sequence A. In practice, since it is considered to be more convenient for the user if the display of the text comes out sooner (the output of speech sound is slightly delayed), this matter can be arbitrarily set.

[0076] Next, a case where the data recording medium is a magnetic tape that cannot be accessed at random is described as an embodiment shown in Fig. 14 and Fig. 15.

[0077] In this embodiment, the basic mode of data recording and reproduction is the same as the above-described embodiments, and reproduction by skipping data on the magnetic tape is not easy. Therefore, it is desirable to provide a buffer for temporarily restoring data in order to make the reproduction unit practical. Concretely, the respective small units corresponding to speech sound data sequences A, B, and C (three tracks) of the above-described virtual tracks are aligned in order so that the corresponding units of speech sound data sequences A, B, and C are put together in this manner:  $A_1, B_1, C_1, A_2, B_2, C_2, A_3, B_3, C_3 \dots$ . Thus, when speech sound data sequence A is reproduced, only the segments of speech sound data sequence A such as  $A_1, A_2, A_3 \dots$  are chosen for the reproduction and when speech sound data sequence B is

reproduced, only the segments of speech sound data sequence B such as  $B_1$ ,  $B_2$ ,  $B_2$  ... are chosen for the reproduction. The same can be carried out for speech sound data sequence C. At this time, in order to reproduce segment  $A_2$  next to segment  $A_1$  without interruption, segments  $B_1$  and  $C_1$  that exist between these segments must be skipped. This is not easy to do with a conventional tape, and therefore data must be brought into the buffer ahead of time from the magnetic tape. Thus, the skipping process and reproduction are carried out in the buffer.

[0078] In the case where this magnetic tape is a so-called DAT (digital audio tape) the following process occurs. First, speech sound of a native speaker recorded on track A is separated into certain units as segments of speech sound data sequence A. For example, the speech sound is separated into units of a constant period of time such as one second. On the other hand, track B (speech sound data sequence B) records the contents spoken slowly (for example, at  $1/3$  the speed of track A on average). Regions are secured so that if a region of 1 second is taken up on track A, it corresponds to a region of four seconds on track B. In addition, the regions are secured so that a region of 9.4 seconds can be taken up on track C for the explanation. The above described number of seconds here

is an example and the present invention is not limited to this. In this example, 14.4 seconds are used as one unit. That is to say, three segments of tracks A, B, and C (equivalent to the respective speech sound data sequences A, B, and C) that correspond to each other are put together to make up 14.4 seconds. It should be noted that a portion of one second on track A is equivalent to one second that is cut out of the original sound data. A portion of four seconds on track B may have a content that is within four seconds, but may finish in 3 seconds, for example, with an extra second remaining. Here, a method for skipping this remaining portion (a one second portion) at the time of reproduction is described below.

[0079] A case of a rotary head type DAT is cited as follows. Namely, a system wherein two tracks are recorded or reproduced during one rotation of the rotary head is generally used in this rotary head type DAT. Thus, speech sound data of 2,880 bytes can be recorded on one track. Accordingly, in the case where one segment is formed of 30 tracks the segment is made up of 86,400 bytes. In the case where the speech sound utilized in the above described embodiments is recorded in this segment at a sampling rate of 48 (Kbit/sec),  $86,400 \text{ (bytes)} \div 6,000 \text{ (bytes)} = 14.4 \text{ (sec)}$  can be recorded.

[0080] Next, the buffer is described.

[0081] When a RAM of 1 MB is utilized, 12 buffers can be provided as a result of this equation:  $10,480,576 + 86,400 \div 12,135$ . The 12 buffers are arranged to function in a ring form, so that two buffers are always selected for a segment data transfer from the tape to the buffers. Reproduction is carried out by utilizing the data transferred to these buffers. When the reproduction is stopped at an arbitrary place, the speech sound (already emitted from the speaker as a reproduced sound) before the point of time where the reproduction is stopped remains in ten buffers. That is to say, speech sound for ten seconds of track A can be repeatedly reproduced from the buffers without rewinding the tape.

[0082] In such a configuration, the 10 seconds of speech sound data that remains in the ten buffers can be easily transferred to any of the virtual tracks A, B or C in an arbitrary segment in the same manner as the above described embodiment using a CD-ROM.

[0083] Here, random access cannot be carried out like a CD-ROM and, therefore, the total period of time of the segments of the three tracks A, B and C that correspond to each other must always be within 14.4 seconds. At this time, the portions of virtual track A may be mechanically separated into constant time intervals (for example, one second intervals). The

portions of track B may be two seconds or five seconds, for example, and therefore variable. A period of time gained by subtracting periods of time of tracks A and B from 14.4 seconds is the period of time that is allocated for track C. Here, it is not necessary for both tracks B and C to record the portions that completely correspond to the portion in the segment of track A, but rather may form a coordinated unit of a number of segments in the vicinity.

[0084] A DAT has a sub-code region and therefore data of byte numbers indicating the respective segments (for every segment, 14.4 seconds/30 tracks) and interfaces between tracks A, B and C in the respective segments can be recorded in this sub-code region.

[0085] Next, an amount of speech sound data that can be recorded on a DAT is concretely calculated.

[0086] In general, a DAT of one hour can record 240,000 tracks. One segment is made up of 30 tracks and the following 6 tracks are used as pause tracks so that the portions recording speech sound data do not become scratched even in the case where the rotary head passes through a great number of times when the tape is stopped. In such a configuration one unit is formed of 36 tracks so that 6,666 units can be recorded in total which becomes  $6,666 \times 14.4 \text{ seconds} = 126 \text{ hours, } 39 \text{ minutes and } 50$

seconds. One hour out of that is allocated for track A and therefore 25 hours, 39 minutes and 50 seconds can be used for tracks B and C. Even if four hours are used for track B spoken slowly, more than 21 hours can be allocated for track C for the explanation and, therefore, the tape has a sufficient length.

[0087] Here, as for how the segments of tracks A, B and C are created, track A may be divided at places where there are some pauses according to speech sound, so that the length of the segments become approximately one second, instead of dividing track A into constant intervals of one second in the same manner as in the above described application of a CD-ROM. That is to say, the length of the segments may vary according to the edit policy at the time of recording. Accordingly, a medium that can be randomly accessed, such as a CD-ROM, may have segments created by a separation into constant time intervals.

[0088] Next, a description will be given of a reproducing method of a speech sound recording medium according to the present invention, in particular, an interrupt handling concerning a repeated reproduction request and switching reproduction by use of Fig. 16 through Fig. 18.

[0089] Fig. 16 is a time chart for explaining an operation to determine an interruption-target segment in a reproducing



method of a speech sound recording medium according to the present invention.

[0090] In general, in a reproducing method of a speech sound recording medium according to the present invention, segments (... , N-1, N, N+1, ...) of a previously instructed speech sound data sequence are reproduced in order. When an interrupt request (shown by R1, R2, and R3, ... in the drawing) concerning a repeated reproduction, switching reproduction, etc., is generated from a learner (for example, an English learner, etc.,) during reproduction, first, a segment of a target for the interrupt request is specified. Namely, as shown in Fig. 16, when segment N is being reproduced, whether such an interrupt request was before or after a boundary position  $S_N$  indicated by boundary position identification data recorded on a header portion of each segment described in the foregoing is judged. Here, in the drawing, boundary positions in the segments are shown by ...,  $S_{N-1}$ ,  $S_N$ ,  $S_{N+1}$ , ..., respectively. Furthermore, based on a result of judgement as to generation timing of this interrupt request, the segment N during reproduction or the previous segment N-1 is specified as a target segment of the generated interrupt request. By specification of a target segment of the generated reproduction request, a reproducing operation of speech sound data is

restarted, according to the contents of instruction of the interrupt request, from the head of the specified segment or the head of a segment of another speech sound data sequence corresponding to the specified segment.

[0091] In detail, in a case where a desirable interrupt request R1 is generated during reproduction of a segment N of a certain speech sound data sequence, for example, generation timing of this interrupt request R1 is generation in a first half part of the segment N. In this case, it is judged to be an instruction delay of an operator, and a previous segment N-1 of the segment N during reproduction is specified as a target segment of the interrupt request. On the other hand, in a case where an interrupt request R2 is generated in a second half part of the segment N during reproduction, this segment N itself during reproduction is specified as a target segment of the interrupt request. Herein, similarly, in a case of an interrupt request R3 generated during reproduction of another segment, for example, a segment N+1 shown in Fig. 16 as well, generation timing of an interrupt request is judged on the basis of a boundary position  $S_{N+1}$ , whereby a target segment of the interrupt request is specified.

[0092] Fig. 17 is a flowchart for explaining a repeated reproduction interrupt handling (a first embodiment according

to a reproducing method) in a reproducing method of a speech sound recording medium according to the present invention.

[0093] It is sufficient that, in a speech sound reproducing medium prepared by this reproducing method, at least, a speech sound data sequence, which is a speech sound data sequence corresponding to one or more sentences made up of a plurality of word sequences to be reproduced and outputted by a predetermined speech sound reproducing means, sectioned into variable-length segments for each of speech sound data respectively divided at pauses in speech sound, and for each section of the speech sound data sequence, boundary position identification data indicating a boundary between a first half part and a second half part of each section split according to the state of speech sound are recorded in desirable regions, respectively.

[0094] First, during reproduction of a speech sound recording medium as in the above, when an interrupt request concerning a repeated reproduction is generated, data of a segment during reproduction (the segment number) is temporarily retained by a predetermined retaining means (step ST1), and data concerning a generating position of the interrupt request is also temporarily retained (step ST2). Then, a boundary position that has been recorded on a header portion of the segment during

reproduction and the interruption generating position are compared along the time base (step ST3). At this time, if the interrupt request is generated after the boundary position, the segment itself during reproduction is specified as a target segment of the interrupt request, and head position data of the specified segment is read out from the directory, and a speech sound reproducing operation is restarted (step ST4). On the other hand, if the interrupt request is generated before the boundary position, a previous segment of the segment during reproduction is specified as a target segment of the interrupt request, and head position data of the specified segment is read out from the directory, and a speech sound reproducing operation is restarted (step ST5).

[0095] Here, in this embodiment, in an instructed repeated reproducing operation, a specified segment's speech sound data is repeatedly reproduced until a next, new instruction is made.

[0096] Next, a switching reproduction interrupt handling (a second embodiment according to a reproducing method) in a reproducing method of a speech sound recording medium according to the present invention will be described by use of a flowchart of Fig. 18.

[0097] It is sufficient that, on a speech sound reproducing medium prepared by this embodiment, at least, a speech sound

data sequence A, which is a speech sound data sequence corresponding to one or more sentences made up of a plurality of word sequences to be reproduced and outputted by a predetermined speech sound reproducing means, sectioned into variable-length segments for each of speech sound data respectively divided at pauses in speech sound; a speech sound data sequence B, which is a speech sound data sequence equivalent to this speech sound data sequence A and made up of slowed speech sound, sectioned into variable-length segments for each of speech sound data equivalent to speech sound data of each segment of the speech sound data sequence A; in order to reproduce the speech sound data sequences A and B by switching, recording position identification data indicating each switchable segment by a recording position of each segment in the recording medium; and for each segment of the speech sound data sequences A and B, boundary position identification data indicating a boundary between a first half part and a second half part split according to the state of speech sound (judged to be cohesiveness of speech sound) are recorded.

[0098] First, during reproduction of a speech sound recording medium as in the above, when an interrupt request concerning a switching reproduction is generated, data of a segment during

reproduction (the segment number) is temporarily retained by a predetermined retaining means (step ST6), and data of a newly instructed speech sound data sequence is retained (step ST7). In addition, data concerning a generating position of the interrupt request is also temporarily retained (step ST8). Then, a boundary position that has been recorded on a header portion of the segment during reproduction and the interruption generating position are compared along the time base (step ST9). At this time, if the interrupt request is generated after the boundary position, the segment itself during reproduction is specified as a target segment of the interrupt request, and head position data of the specified segment, which is a segment of a speech sound data sequence specified by the interrupt request, is read out from the directory, and a speech sound reproducing operation is restarted (step ST10). On the other hand, if the interrupt request is generated before the boundary position, a previous segment of the segment during reproduction is specified as a target segment of the interrupt request, and head position data of the specified segment, which is a segment of a speech sound data sequence specified by the interrupt request, is read out from the directory, and a speech sound reproducing operation is restarted (step ST11).

[0099] Herein, the present invention is not limited to the

above-described embodiments but rather a variety of modifications are possible.

[0100] The present invention can be applied to a system, for example, that connects to a video disk or a video tape. That is to say, a soundtrack of a movie is placed on track A spoken at regular speed (speed of a native speaker); speech sound spoken slowly is placed on track B; and Japanese explanation is placed on track C. Thus, the reproduction of track A can be stopped at a place that cannot be understood and slightly moved back (at this time, the image can be held while the system can be practically used) so that track B can be heard and if it cannot still be understood track C can be used. Here, the speech sound may be synchronized with the images only at the time track A is reproduced.

[0101] In addition, the present invention can be used on a personal computer or the like. That is to say, a high level application becomes possible by combining with a system connecting a CAI (Computer Aided Instruction) or a programmable high level system such as a CDI (Compact Disk Interactive). In addition, it is also possible to use the memory device of a personal computer in place of the above described CD or DAT in order to implement the present invention.

[0102] Furthermore, the present invention can be applied to

the study of recitation of Chinese poems, the study of law, in addition to the study of English conversation. Moreover, data sequences are not limited to three types or four types, but rather may be no less than these types.

[0103]

[Effects of the Invention] As described above, according to the present invention, a speech sound data recording medium on which, together with various types of speech sound data, as data concerning variable-length segments to make up the speech sound data sequence, at least boundary position identification data indicating a boundary between a first half part and a second half part split according to the state of speech sound of each segment in the recording medium has been recorded; whether timing of a predetermined interrupt request is before or after a boundary position of a segment during reproduction is judged; and based on a result of the judgement, a target segment of the interrupt request is specified, therefore, an effect exists such that flexible reproducing operation with an operator's instruction delay taken into consideration can be realized.

[BRIEF DESCRIPTION OF THE DRAWINGS]

[Fig. 1] Views for conceptually explaining various types of data including speech sound data sequences to be recorded on



a speech sound recording medium according to the present invention.

[Fig. 2] A view for explaining each speech sound data sequence of a speech sound recording medium (third embodiment), according to the present invention, applied to self-study English conversation and recording contents thereof.

[Fig. 3] A table for explaining, concerning each speech sound data sequence shown in Fig. 2, a relationship between time and capacity of each segment.

[Fig. 4] A table for explaining the recording contents (including recorded position identification data) of a directory region in the speech sound recording medium shown in Fig. 2 and Fig. 3.

[Fig. 5] Views showing a variable-length segment (including boundary position identification data) to be recorded on a speech sound recording medium according to the present invention.

[Fig. 6] A perspective view showing an overall construction of a reproduction unit to realize a reproducing method of a speech sound recording medium according to the present invention.

[Fig. 7] A block diagram showing a construction of the reproduction unit shown in Fig. 6.

[Fig. 8] Views for explaining a condition of memory assignment to the ROM and RAM shown in Fig. 7.

[Fig. 9] A plan view showing an overall construction of the handset and earphone shown in Fig. 6.

[Fig. 10] Diagrams for explaining a basic reproducing sequence of a reproducing method for a speech sound recording medium according to the present invention.

[Fig. 11] A view showing another construction example of a directory region in a speech sound recording medium according to the present invention.

[Fig. 12] A view for explaining a condition of segmentation of each speech sound data sequence corresponding to recording contents of the directory region shown in Fig. 11.

[Fig. 13] A view for explaining, in a divided condition shown in Fig. 12, a relationship between the values of the second and third bits of attribute C which has 8 bits of recording contents of the directory region and the divided respective segments.

[Fig. 14] A view showing another construction example of a directory region in a speech sound recording medium according to the present invention.

[Fig. 15] A plan view showing a schematic construction of a handset of a reproduction unit to reproduce a speech sound

recording medium having the directory shown in Fig. 12.

[Fig. 16] A diagram (time chart) for explaining an operation to determine an interruption-target segment in a reproducing method of a speech sound recording medium according to the present invention.

[Fig. 17] A flowchart for explaining an interrupt handling concerning a repeated reproduction in a reproducing method of a speech sound recording medium according to the present invention.

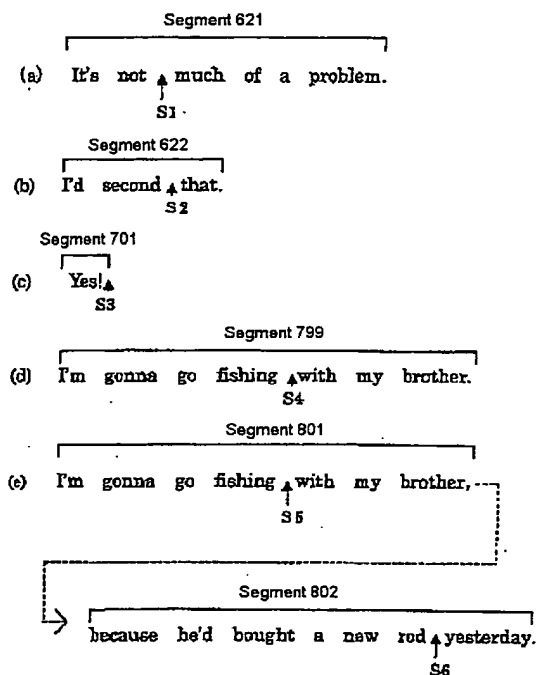
[Fig. 18] A flowchart for explaining an interrupt handling concerning a switching reproduction in a reproducing method of a speech sound recording medium according to the present invention.

[Fig. 19] A diagram (time chart) for explaining a repeating interrupt operation in a conventional reproducing method of a speech sound recording medium.

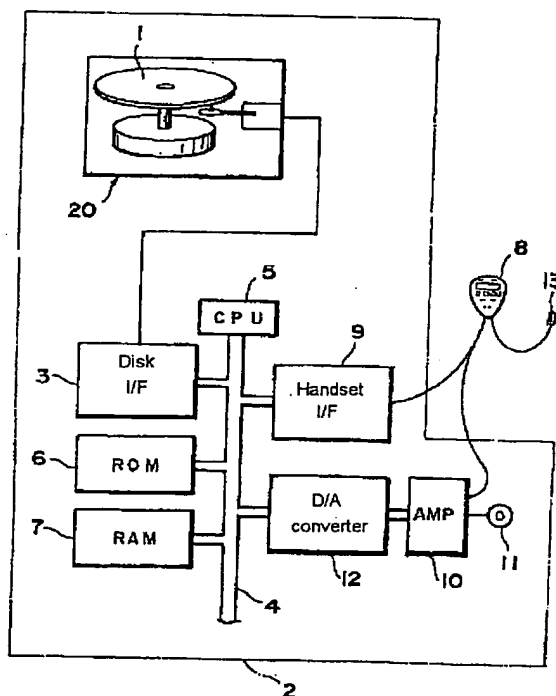
[Description of Symbols]

1 ... Recording medium, 2 ... Reproduction unit, 8 ... Handset, 13 ... Earphone, 21, 210 ... Display part, 23A, 23B, 23C, 23D ... Reproducing data sequence instruction switch, 25A ... Repeat instruction switch

[Fig. 1]



[Fig. 7]



[Fig. 2]

Recorded Contents		
A	"It's not much of a problem."	I'd second that."
	621	622
B	"It is not much of a problem."	I would second that."
	621	622
C	(621)	"'it' indicates the previous 'something' and 'not much of a problem' is an idiom meaning 'no problem' or 'no worries'.
	(622)	Next, 'I'd' is an abbreviation of 'I would' and 'I could' can also be abbreviated to 'I'd' in the same manner. 'Second' means to 'support' or to 'agree with' and 'I'd second that' means 'I agree with that'."

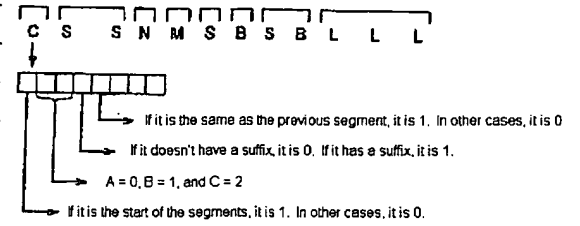
[Fig. 13]

Second or Third Bit	Segment
0	621
1	621-1
1	621-2
2	621-1
2	621-2
2	621-3
2	621-4
0	622
1	622
2	622-1
2	622-2

[Fig. 3]

Segment			621	622	
Time (Second)	A		2.0 (0.3)	1.3 (0.4)	
	B		3.9 (1.1)	2.6 (1.4)	
	C		12	25	
Capacity (KB)	A		12	7.8	
	B		23.4	15.6	
	C		72	160	

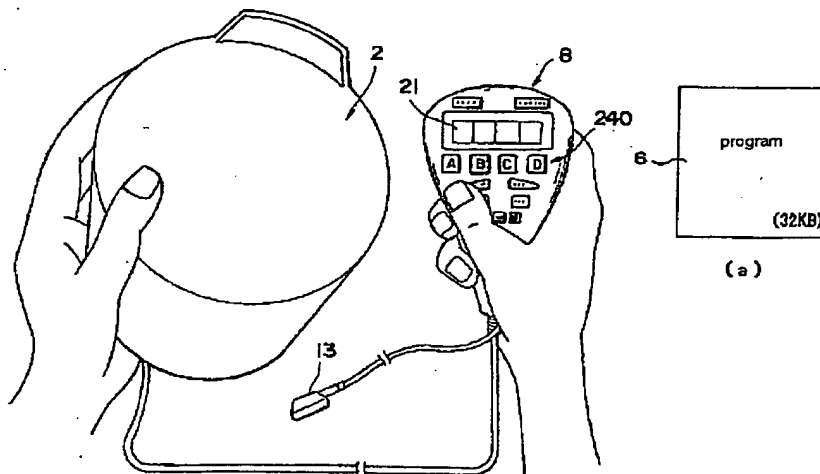
[Fig. 11]



[Fig. 4]

Data sequence	Segment	Category	Positional data				Start byte	Quantity of segments (Byte)
	SS	C	M	S	B		SB	LLL
A								
A	621	0	0	11	3		826	6,000
B		64	0	11	3		2,026	17,400
C		128	0	11	6		1,282	72,000
A	622	0	0	11	49		2,018	4,800
B		64	0	11	50		570	12,600
C		128	0	11	52		1,274	150,000

[Fig. 6]

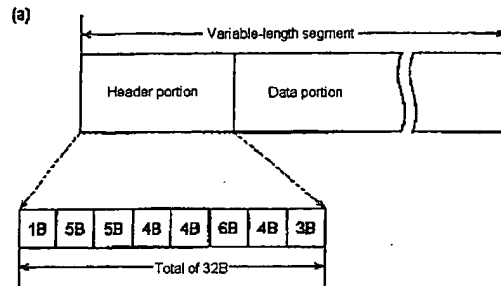


[Fig. 8]

For track	(50KB)
Double buffer	(50KB)
For directory	(75KB)
For directory	(75KB)
System area	(8KB)

(b)

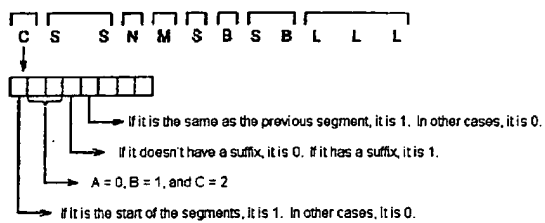
[Fig. 5]



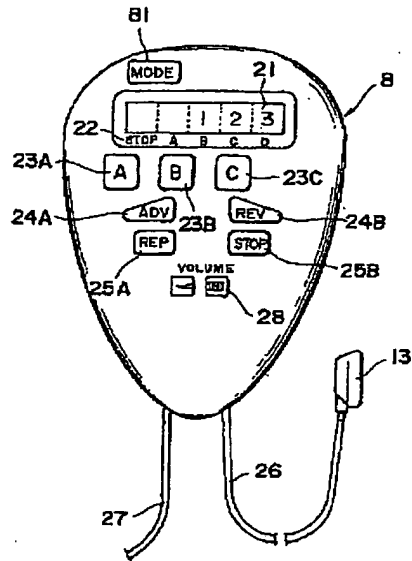
(b)

Header (Total of 32B)	Data sequence	Contents
1B		Presence of text data and image data, etc. (1B)
5B	A	Data sequence type (1B), Data length (3B), Boundary position (1B)
5B	B	Data sequence type (1B), Data length (3B), Boundary position (1B)
4B	C	Data sequence type (1B), Data length (3B)
4B	(D)	Data sequence type (1B), Data length (3B)
6B	D	Address (3B), Data length (3B)
4B	E	
3B	Reserve	

[Fig. 14]

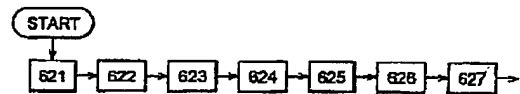


[Fig. 9]

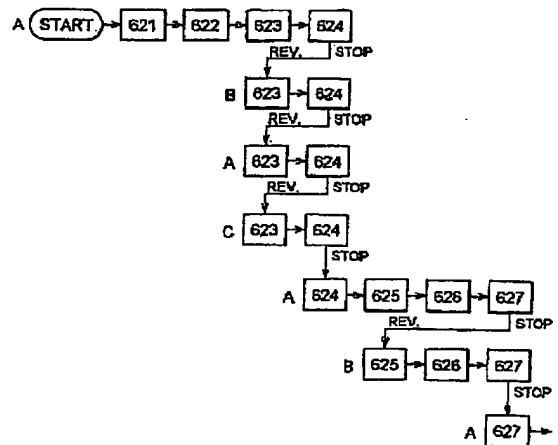


[Fig. 10]

(a)



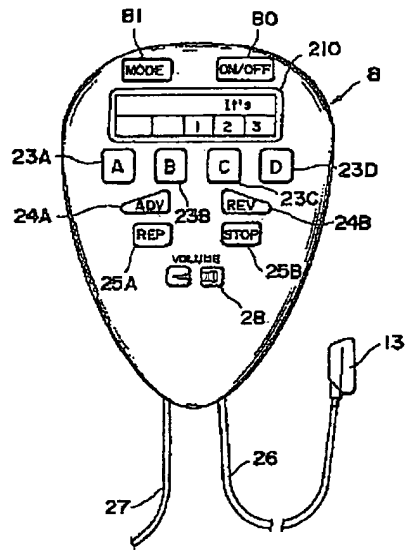
(b)



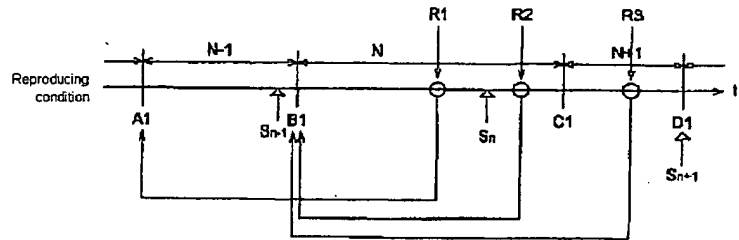
[Fig. 12]

Data sequence	Segmentation
A	<div> <div>It's not much of a problem.</div> <div>I'd second that.</div> </div> <div> <div>621</div> <div>622</div> </div>
B	<div> <div>It's not much of a problem.</div> <div>I'd second that.</div> </div> <div> <div>621-1</div> <div>621-2</div> <div>622</div> </div>
C	<div> <div>621-1</div> <div>621-2</div> <div>621-3</div> <div>621-4</div> <div>622-1</div> <div>622-2</div> </div>

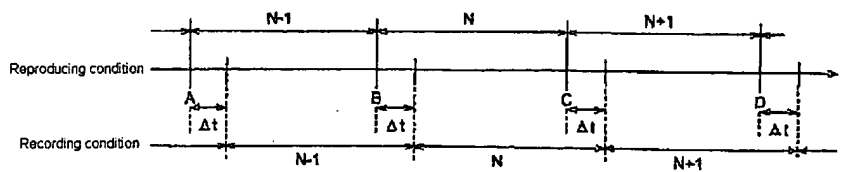
[Fig. 15]



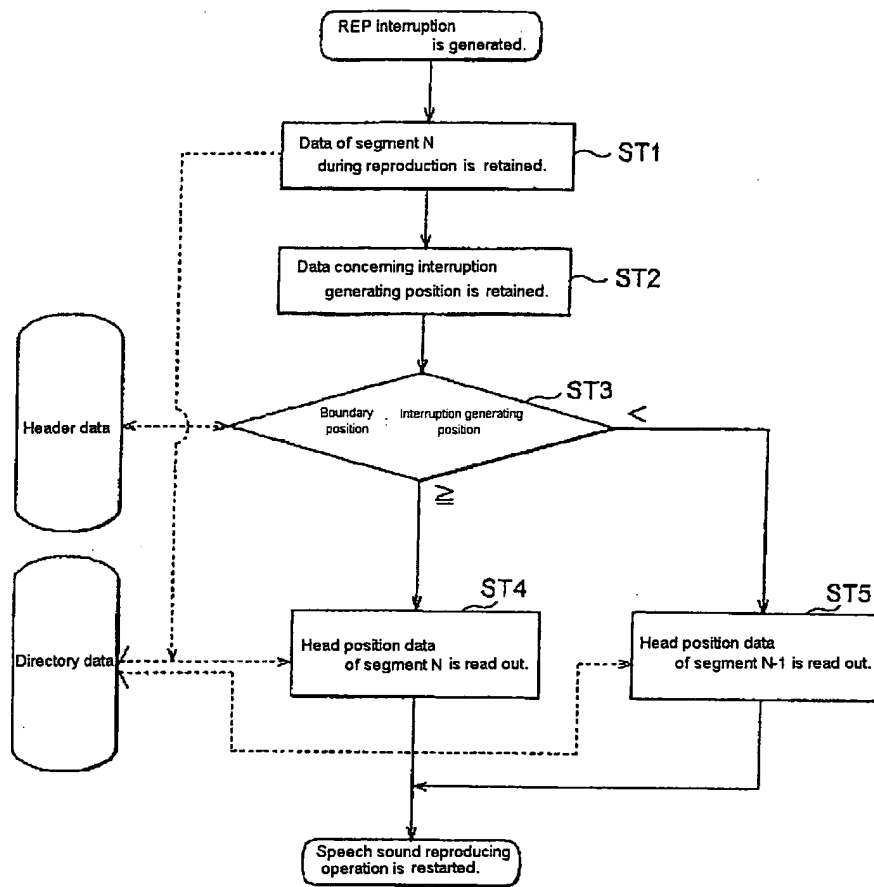
[Fig. 16]



[Fig. 19]



[Fig. 17]





[Fig. 18]

